The Practical Value Aspect of Semantically Structured Information Environments

Investigating the implied logical reasoning potential of semantic layers through the design of persuasive recommender systems for data-driven businesses

Master's Thesis by

Joachim August Overgaard
Nicklas Thorup Poulsen

The Faculty of Humanities

Department of Communication & Psychology

Information Architecture and Persuasive Design

BREAKING NEW GROUND
AALBORG UNIVERSITY

Title Page

Title: The Practical Value Aspect of Semantically Structured Information Environments

Study: Information Architecture and Persuasive Design

Semester: 10th semester

Semester theme: Master's thesis

Specialization of master's degree: Master of Science in IT (cand.it)

Institution: Aalborg University

ECTS points: 30

Supervisor 1: David Jakobsen

Supervisor 2: Christian Grund

Student 1: Joachim August Overgaard

Student 2: Nicklas Thorup Poulsen

Pages: 177

Normal pages: 116,04

Characters: 278.503

Appendices: 3

Submission date: 31-05-2021

Table of Content

Table of Figures

Table of Tables

1 Acknowledgements

This thesis has been a challenge to construct in the midst of the corona-pandemic. Therefore, a very heartfelt *thank you* to friends and families for continuous support and a means of social relief in these stressful times.

A big thank you to our initial supervisor *David Jakobsen* who saw the potential in our thesis right away and motivated us to work hard in reaching our end goal. Likewise, we would like to acknowledge *Christian Grund* for standing in as supervisor, providing the last pieces of guidance that was needed in order to hand in properly.

We greatly appreciate our collaborators in OpdagDanmark, *Lau Starcke-Jensen* and *Sebastian Søndergaard*. We also highly cherish the help from the Communications and Marketing-department in NNIT, *Lars B. Petersen*, *Birgitte Barslund*, and *Peter N. Jansen*. A big thanks to all of you for giving us the possibility of working with your respective businesses as cases and providing relevant insights into your information environments.

Lastly, we would also like to give our special thanks to *Matthew Ives*, for the incredible help he has provided in terms of grammatical inspection. We are very grateful for your support.

2 Abstract

Although many data-driven businesses sufficiently utilize data in their practices, they often overlook the opportunity to structure data semantically, as they are not aware of the value potential. This thesis investigates the potential value of establishing a semantic layer in business structures and how this can consequently improve the recognition of this aspect of information architecture in professional environments. To showcase the instrumental value aspect of semantically structured data, a concept consisting of a design process for a persuasive recommender system will be utilized as a tangible means of reference. The system will incorporate formal ontology and persuasive principles in a targeted information environment in order to argue for both the most logical content recommendations for each specific domain user, but also a more persuasive and risk-aware means of communication. This highlights some of the practical value potential that can be found through semantic recommendation, while simultaneously presenting the opportunity to obtain new logical insights from targeted information.

3 Introduction

Information architecture is an academic field which incorporates many different elements in regard to dissemination of the broad concept of digital information. For example, how to present information to make it more accessible, easier to navigate or make certain information stand out etc. This can be in relation to sorting massive amounts of information, but also when it comes to changing perspectives by rearranging pieces of information to construct arguments in new ways. The subjects that this field particularly touches upon are, for instance, persuasive design, formal ontologies, taxonomies, navigation-systems, search-algorithms and of course the ethics that surround these areas. However, this is just the tip of the iceberg in regard to practical functions within the field, as there is a vast array of informational aspects that the field can help to monitor.

In order to understand the potential of information architecture, information itself must first be understood. When talking about the architecture of information, it is always in relation to information environments. Information environments are any space/place where one is able to learn and understand the concept of things. It is a space where we communicate with each other in one way or another (Arango, 2018). Within these information environments, you have the kind of information architecture that is commonly thought of when discussing practical information architecture in organizational matters - the structures of websites, intranets, applications etc. These are all overarching functionalities where the structural architecture and purpose of the information environment are clearly understood based on their visual nature.

One kind of information architecture is, however, often neglected when discussing its practicality in organizational matters. The semantic potential of the deeper layer of available information, that may be present in an information environment, is rarely deliberated for practical purposes (Tesfaye, 2020). In an information environment there might exist knowledge that has yet to be structured or utilized properly. These can be the result of data/information that has yet to be collected. It can also be the result of having collected data from disparate sources with no structural architecture.

Both of these cases are examples of an information environment with deeper knowledge and potential value that is not obtained. Many modern companies understand the value of working with the structural architecture of their front-end solutions. Creating an information environment which must present the company and its products in a professional manner (a website) has a clear and understandable business value. However, the business value of a semantically structured information environment is much harder to grasp. It is still a rarity to find companies that fully grasp the semantic potential of their data. One of the responsibilities of information architects is to assist people or companies in structuring, understanding and utilizing their semantic potentials with data. This can be accomplished by implementing what is known as a semantic layer (Tesfaye, 2020).

A semantic layer bridges complex data with front-end solutions like websites, applications etc. (Figure 1). The layer makes understanding the full business potential of a given concept (front-end solution) tangible by mapping different sources and channels of related data, creating a unified model that is easy to understand for experts and novices alike. A semantic layer is a combination of organizational data, data models, semantic structures, and tools that enable implementation and up-scaling. Through these aspects, it provides new insights and uses for data, showing increased value of the given concept in the process. With a semantic layer, the potential business value that is to be gained can easily be presented and understood by companies.

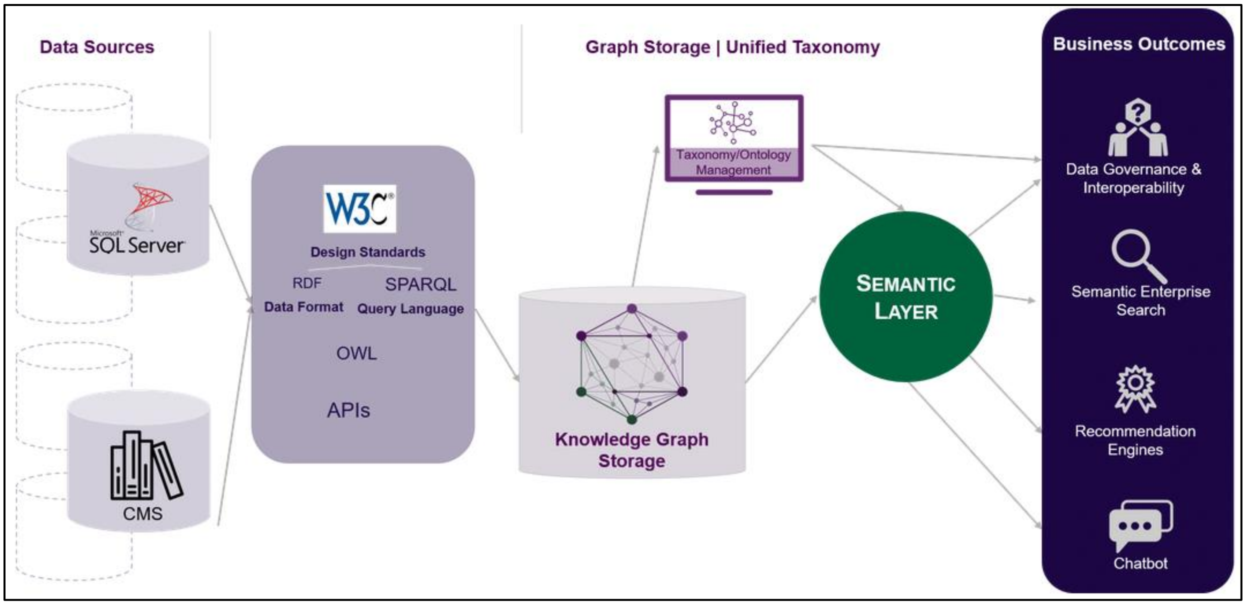

Figure 1: The semantic layer's relation to data sources and business outcomes (Tesfaye, 2020).

With this understanding of information environments and semantic layers, there is a clear problematic direction. Many companies succeed in working with information architecture related to the structuring and set up of their solutions and services. However, they fail to obtain the full value of their information environments as a consequence of neglecting semantically structured data. This implication leads to a potential for information architects to showcase the business value of this field. As such, this thesis attempts to present the practical applicability and value of information architecture regarding the development of semantic layers in organizational information environments. The topic of semantic layers is beginning to become more relevant, as expanding data collection practices and opportunities emerge. As data-driven businesses are starting to get a hold of ever larger amounts of information, competently developed information architectures are becoming more valuable.

An example of this would be the case of the web-based media company, Pinterest, that sought counseling for the large amount of information that they were storing in their system. Information architects helped place their data in a new taxonomy which was furthermore constructed as a formal ontology (Gonçalves et al., 2019). This was done with the help of Protégé (an OWL editor tool), which made it extremely easy for the existing employees at Pinterest to utilize, oversee, and most importantly, further develop the finalized ontology. This proved to be a very valuable investment for Pinterest and strengthened not only the economical aspect of their business, but their overall perspective on data-collection and the proper utilization of their resources as well.

There is a big difference in the value that can be gained between trying to find patterns from collected raw data, and creating properly structured systems for collected data, which make it evident how to utilize the data properly. The Pinterest-case is an example of a company gaining value from realizing the potential of semantically structured data. Like with most modern companies, Pinterest were already utilizing basic excel sheets of syntax data before they delved into assimilating their information semantically in an ontology (Gonçalves et al., 2019). Working with, and evaluating, their information architecture helped them significantly with establishing an understanding of what their data was capable of in a broader perspective. This case is likely one of many to come for businesses on the market today. To these companies, their current data and data practices might seem adequate or even optimized fully, but the reality is different. With semantic layers, they have the potential to derive meaning and interoperability of data, improve their processing of data, unify data across business domains through governance, and provide definitions of data to machines (Tesfaye, 2020).

Data, especially data about users of certain products (big data), has become a more important resource in the last couple of years. It is, for instance, utilized in businesses to improve their existing products and services or help advertisers with ad-targeting for their campaigns. In this regard, data is starting to become known as “the new oil” based on the sheer importance of data worldwide, and the fact that it now has consequences for many people around the world (Bhageshpur, 2019). Data is currently problematized on the basis of the public debate surrounding online privacy. Data is what makes it possible for big corporations to sell information about the average user of their product to the highest bidder, whoever that may be. This, however, is only one side of what data is capable of in the grand scheme of things.

Utilizing the semantic layer alongside data, and thus a more complex and holistic take on the overall information potential, can be an important business decision in order to not lag behind (Miller, 2019). This makes the information architect more relevant as a profession, and has the potential to spark interest in hiring people from the field across many different types of data-driven businesses. Furthermore, there is an academic angle concerning the practical value of semantic layers, built through selected information architecture methods and theories. This brings forth an interesting examination opportunity that can establish a starting point for this thesis.

In order to derive scientific insights from the potential of the practical value that can be gained by incorporating semantic layers in data-driven businesses' information environments, the scope of this thesis needs to be manageable. Therefore, it is important to determine a narrow field of optimization to focus on. Choosing an area of interest that is both meaningful for most data-driven businesses and relevant for information architects from an academic and practical perspective. This will strengthen the thesis' relevance. Therefore, the area of interest is determined to specifically concern the concept: "*to recommend content*", based on the relevant data that a business possesses. Because of this, the companies that will be chosen are those with either existing data sources or a strong potential to acquire relevant data, which also have an interest in presenting content to their users in new and amplified ways. Amplified in the sense, that users receive quality content recommendations that they find relevant and valuable in a higher regard. Furthermore, this may increase the users' satisfaction in the product or the image/brand of the companies in question.

Recommendation of content is often an important factor in the relationship between users and companies for increasing the users' interest in the company's content. For this thesis, it is also a way to showcase the potential that lies in a semantic layer. From a business perspective, recommendations can save money and increase customer retention by increasing the conversion rate, selling more diverse items, or improve user satisfaction, for example ("5 Best Practices for Effective Personalized Product Recommendations", n.d.). However, the actual earnings, gained as a result of a recommendation service, might be challenging to demonstrate depending on the service. Recommendation is often tied to the overall design of the user experience which shows profit in a different, albeit more lasting way. Unless the recommendation service is already in place, and its recommendations are traceable through defined parameters of success, its success becomes ambiguous.

Regarding the customer journey, where the customer typically goes through stages like need, awareness, research, comparison, decision and purchase, an intelligent recommender system can assist with every aspect except need and purchase, which establishes the system as essential to the user experience, if implemented the right way (Katakam, 2019). Even though recommendation is the obvious answer for most, it can also be obstructive for some users. Recommendations are therefore not a given for every company. This is why it is worth having the recommendation potential examined by people who are knowledgeable in the targeted domain, to oversee potential problems that might occur in terms of user satisfaction with recommendations.

Successful recommendations are created through effective utilization of data. As such, it is advantageous for information architects to frame and map the entire information environment, in which recommendations are made, by creating a semantic layer. It is important to create an overview of how the different strings of data affect each other in the system, and their logical relationships with each other, as this is what creates the insightful potential for knowledgeable algorithms in a semantic information architecture. For this task, it is advantageous to utilize the formal ontology. Ontologies are conceptual tools utilized to map out all available information inside a closed information domain (Horridge, 2011). Ontologies can be developed to be incredibly simple or advanced, based on the nature of the targeted domain. The formal ontology can, in the context of this thesis, highlight logical insights about new approaches to data dissemination and inferred facts about users that are tied to the data. This gives the companies a clear overview of the logic in a given recommender system, and how the data that is put into the domain affects the possible end results.

The semantic nature of the ontologies offer an alternative to automated machine learning principles, where categorizing information and creating logical conclusions is essentially automated by an algorithm that has been fed a sufficient amount of structured data. This has, in recent times, become a problematic way to utilize data for things like AI-functionality, since it often creates biases with unprecedented solutions. Semantic architecture is therefore a key component in the concept of explainable AI, which makes it very relevant in modern times (Bianchi, Rossiello, Costabello, Palmonari & Minervini, 2020). The semantic nature of explicitly stating logical relations makes the formal ontology act as a transparent framework to create a holistic understanding of information environments, where interoperability is a given.

This is one of the strengths of semantic architecture, and with the imminent integration of AI into current technology, it is an important aspect to factor into the overall value of information architecture. As a matter of fact, modern recommender systems are considered by some to be a weak form of AI, even though they are not tied to the general (perhaps sci-fi) understanding of the term (Bostrom, 2016). This is because the overall defining term for AI is ambiguous at this point in time. However, there is no doubt that we are fast-approaching more nuanced ways that AI can play a role in everyday technology. This also includes recommender systems. A semantic layer is therefore possibly a more ethical approach to future-proofing technology for imminent integration in regard to intelligent automation in coming years. Although this is not specifically a selling point for practicality at this point in time, it is necessary to consider for the long-term utilization of information architecture in conjunction with algorithms such as recommender systems.

In relation to recommendation of content, the process itself can also be investigated. From knowing which criteria creates a good recommendation, to actually having recommended the piece of content that fits the criteria, in other words, to go in-depth with the concept: "*to recommend content*". Expanding upon this conceptual input angle, it is also possible to enhance this aspect with persuasion principles such as B.J. Fogg's Behavior Model (Fogg, 2009) as well as the Kairos-moment (Glud & Jespersen, 2008). These are used to guide and convey the specific piece of content in a way that is persuasive for the user, and assures that the user's relationship to the brand of the business is strengthened, as well as to not compromise the interest that the user feels about interacting with future content.

Exactly how much businesses need formal ontologies and persuasion-principles, in relation to the concept: "*to recommend content*", is not currently known with any degree of certainty. It has been examined that there is a reason to implement it for most content-sharing businesses. However, as mentioned before, recommendations can also annoy certain users. What is known for certain, is that a substantial amount of businesses are struggling with their handling of data (Bean & Davenport, 2019). Still, this does not necessarily prove that they specifically need a semantic layer in their data structure. Data utilized for recommendations is a broad subject, and there are many perspectives regarding its use.

Even though proper data utilization is widely regarded as valuable, the practical value that can be gained differs drastically depending on the company and surrounding contexts. It can depend on other factors than the specific value of the service itself, such as the service's expenses when developing and implementing it for companies. Maybe the service does not have any significant impact on the users because there is no strong basis to keep interacting with content on the platform in the first place. These are factors and contexts that must be considered for any company that invests in data dissemination, and should be included in any investigative work before implementing a recommender system.

All of the above can be summarized as the building blocks for approaching the problem area. However, regarding how the investigation of the problem area will fundamentally commence, it is first and foremost important to state that the problem consists of many businesses not realizing the potential value from utilizing their data semantically. This is at the core of what is assumed to be achievable with the following: that more data-driven businesses invest in their semantic information architecture by developing and implementing semantic layers, resulting in certain increases of value in, for example, better user-experiences (specifically for recommendation engines). This investment in information architecture may even establish premises for employment, as the valuable outcomes of the field become apparent.

With the scope of recommendation engines, principles of information architecture can be utilized to form a concept. The purpose of this concept is to establish a semantic layer. Alongside this semantic layer, a recommender system can be developed to highlight the value aspect of the new information architecture. As a scalable and repeatable concept, semantic layers become easier to develop and implement, making the value of information architecture more prominent to businesses. As such, this thesis will combine an academic investigation with a practical problem-solving perspective applicable to the professional market. The investigative approach to this problem will be further underlined in the hypothesis of this thesis.

4 Hypothesis

Having presented the problem area in the introduction, a hypothesis can be constructed. This is used to properly examine and ensure that the investigative approach is always located at the core of the project. This hypothesis creates a reflective starting point for the further investigative work and will ensure the development of the project as being coherent and meaningful in the long run. The hypothesis is therefore actualized as follows:

"Information architecture is a field which can be highly relevant in assisting data-driven businesses with their use of data. Many data-driven businesses have the potential to improve upon their use of data by deploying a semantic layer to a targeted information environment. Based on these assumptions, the implementation of a user-based content recommendation service, utilizing data handling, formal ontology and persuasion principles, can improve upon the value of a targeted information environment. This will furthermore underline one way in which information architecture can become valuable in professional environments."

Having defined the hypothesis from which the project can base its reasoning and purpose, defining the project's theoretical point of view comes next. In order for the thesis to create scientific knowledge, the *Theory of Science* chapter details what the chosen approaches and methods are, and how they allow for the project to accomplish its purpose as an academic thesis. Even with this theoretical approach, the thesis still retains relevance for stakeholders of targeted businesses in practice.

5 Theory of Science

In order for the thesis to provide academic and valuable knowledge its scientific approach needs to be established. This is accomplished using the following methods and approaches.

5.1 Pragmatic Paradigm

With the project investigating the practical value of a recommender system to showcase the usability for a semantic layer, the thesis needs to focus on the practicality of this sort of information architecture. As such, the thesis is concerned with 'what works' and 'how' for finding solutions to problems that further its overarching goals. For this purpose, the pragmatic paradigm is chosen as the thesis' perspective, ensuring that the capabilities of the thesis are not concerned with specific methods. Rather, creating practical value is emphasized and all approaches that are beneficial for this purpose are considered (Mackenzie & Knipe, 2006). The pragmatic paradigm allows for freedom of choice in regard to methods, techniques, and procedures. This also applies to mixed methods research. Seeing as the pragmatic paradigm has a lot of similarities and beneficial connotations to mixed methods research, this approach to research is chosen for the project (Creswell, 2009).

5.2 Mixed Methods

Similar to pragmatism, mixed methods research does not commit to a specific philosophy, as it draws information from both quantitative and qualitative data. In this project, the intent of the mixed method's research is to allow for the gathering of required quantitative and qualitative data, in order to develop the concept. This will ensure that the project is not limited in its approaches to research and the primary focus is on practicality. Quantitative data is to be collected and interpreted from companies'/clients' platforms and provided data extracts, such as excel sheets of syntax data. This ensures that the recommendation service is built on a base of relevant data, assisting the service in providing competent recommendations. Qualitative data is to be gathered from meetings and general communication with the company/client in question. These meetings will provide the needed contexts and domain expertise which are not present in the quantitative data extractions. These interviews are also likely to reveal information that ensures alignment between the company and the ones responsible for the thesis, ensuring that the overall work goes as planned for both parties (Creswell, 2009). This qualitative means of data gathering is only a means to validate the initial approach to the domain of interest. The meetings are to be understood as collaborative sessions, rather than classical qualitative research interviews.

It is expected that the two types of data will end up building upon each other's findings. Deciding which type of data will come first in the sequential process is not clear, as the sequence may differ depending on the company that is collaborated with. Some companies may have large data sheets that can be delivered quickly, making quantitative data the first in the sequence. Other companies may have to spend a lot of resources and manpower to gather the needed data, making information gathered through meetings and interviews the more immediately available source data. This would make qualitative data the first in the sequence. As such, the sequential process will have to be explored in each case in order to decide on the optimal sequence. With these reflections in mind, the research strategy is chosen to be either the Sequential Explanatory Design or the Sequential Exploratory Design (Creswell, 2009) (Figure 2).

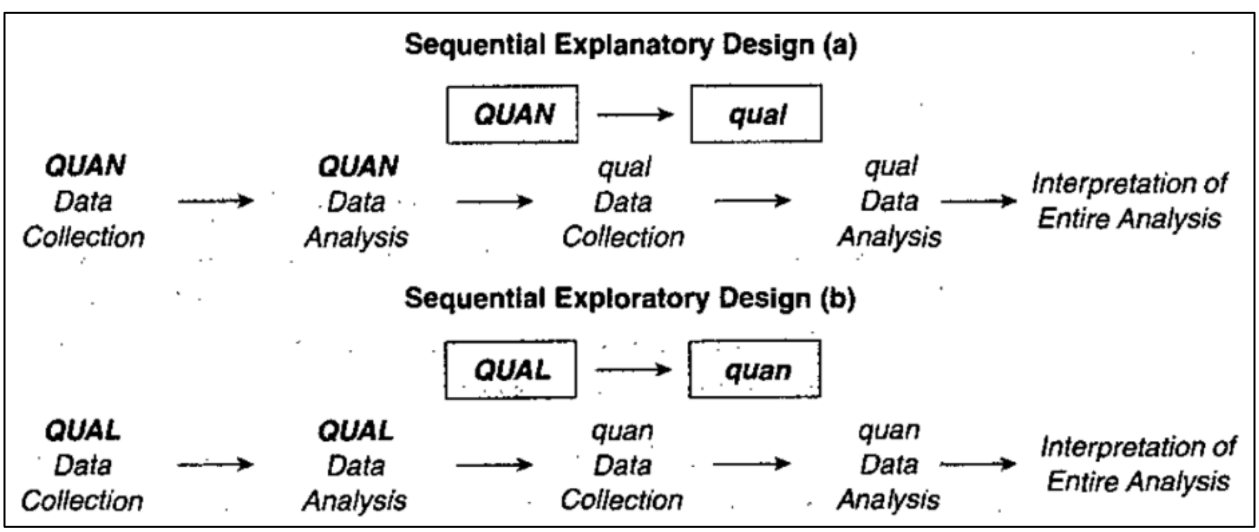

Figure 2: Sequential Explanatory and Exploratory Design (a) & (b) (Creswell, 2009).

5.3 Hermeneutics

The data, and therefore knowledge, that is gained through mixed method research needs a high level of interpretation. For the development of the practical concept, interpretation of data is needed in order to ensure alignment of the concept with the real-life contexts that it is being developed in. Additionally, in order to gain knowledge from the project that is useful and beneficial on an academic level, the thesis demands the use of interpretation when obtaining valuable findings from the insights that the concept provides. As such, Hermeneutics (Theodore, 2020) is introduced to the project, with the Hermeneutic Circle (Alvesson & Sköldberg, 2000) (Figure 3), allowing for new understandings to arise from the exploration of presuppositions through interpretation. The introduction of hermeneutics in the thesis allows for the opportunity to strengthen conclusions and determine the accuracy of findings.

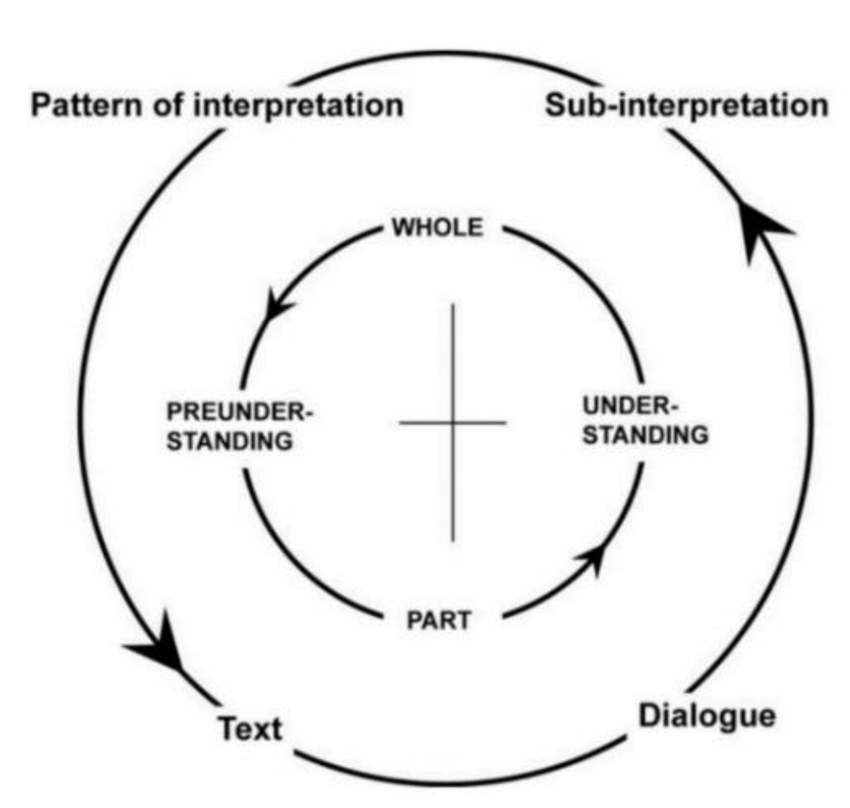

Figure 3: The Hermeneutic Circle (Alvesson & Sköldberg, 2000).

5.5 Approach and Mindset

As the overarching goal of the thesis is to showcase, and conclude on, the practical implications and benefits of utilizing a semantic layer for the development of a recommender system, its design approach is set to be primarily research-led with an expert mindset (Sanders, 2008) (Figure 4). The ones responsible for the thesis, are the candidates that will be able to provide answers to the thesis' success or failure regarding its academic goal to provide answers to the established problem and hypothesis. However, when focusing solely on the practical value that is to be provided through the concept, it could be argued that the design approach mindset should be set as a participatory mindset, rather than an expert mindset (In this thesis, "collaborative mindset" is the preferred term for a participatory mindset, as it implies a more equal collaboration between the parties). Collaboration with the domain experts from a company is a required part of the development of the design for the recommender system. Therefore, the domain experts from the company will also take part in judging the success or failure of the concept's practical value in the specific case.

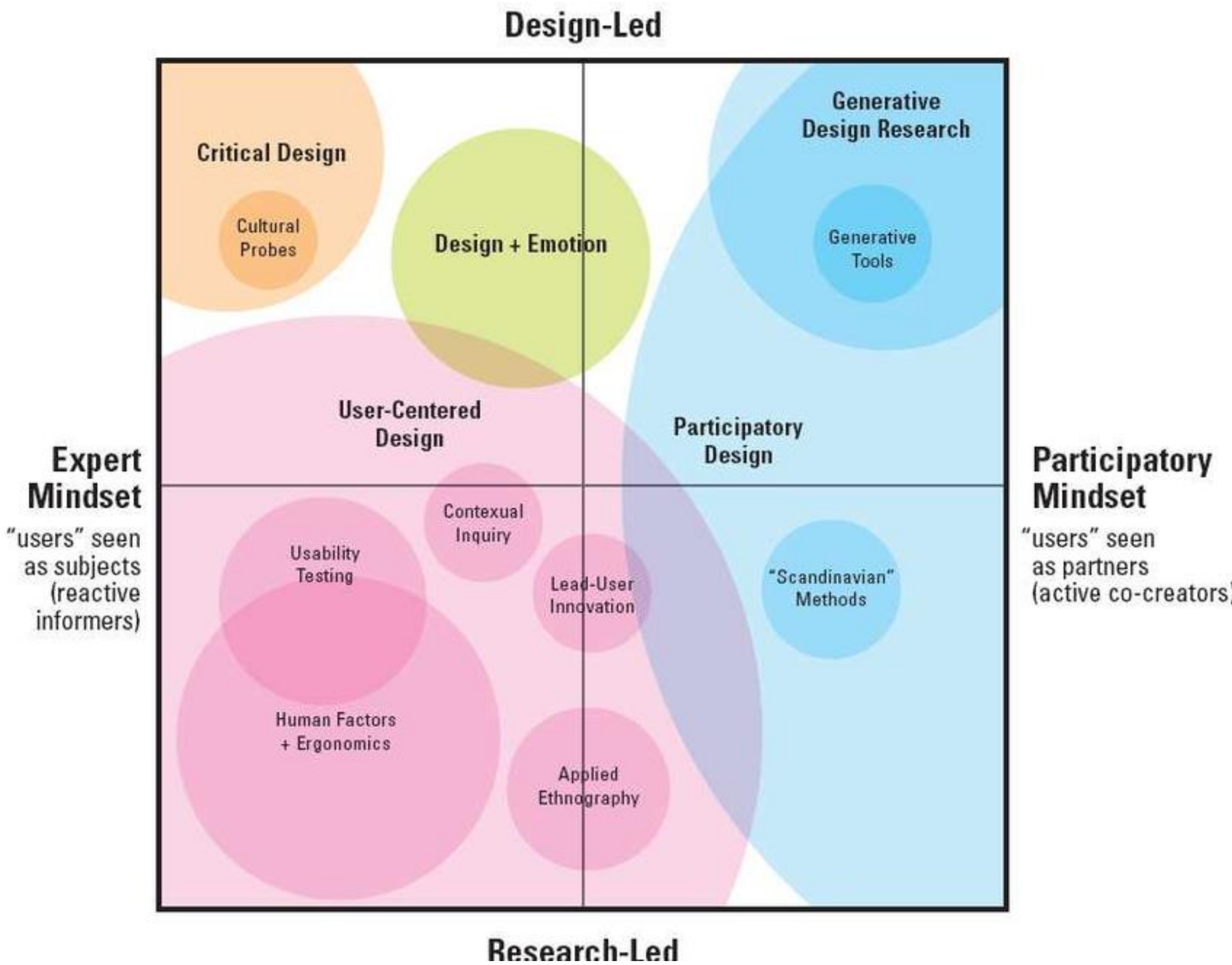

Figure 4: Approaches and Mindset (Sanders, 2008).

Regarding the actual work-process of the concept, the design approach is expected to shift several times. At certain times during the work process, a *collaborative mindset* is expected where the ones responsible for the thesis, and domain experts from the company collaborate. This is also where a design-led approach is expected to be more present, as domain experts (the company) are assumed to have preferences and needs that must be met. At other times in the process, an expert mindset is expected, as it will solely be the ones responsible for the thesis that make the qualified decisions.

6 Concept Introduction

An important aspect of investigating a practical perspective for implementing a semantic layer is, for this thesis, the recommendation of content. This will be demonstrated with an actual market solution, which will be referred to as the concept. Relevant data-driven businesses can utilize this concept to strengthen their content recommendations based on information about their context, content, and users. This solution is to be offered by information architects functioning as consultants, who can, first of all, help businesses establish proper tracking of their data (if they have not), but most importantly assist in getting a clearer view of their data landscape and become aware of new logical perspectives. This can also lead to guidance in terms of how to utilize these perspectives persuasively in their recommendation of content. This ultimately creates a concept which gets businesses up to speed with their data usage, and in this case guarantees insights into how they should recommend their content in the specified information environment. All of this ties into the greater academic purpose by researching the theoretical value potential of such a concept for businesses that implement semantic layers, which ultimately highlights how utilizing a semantic layer creates practical value.

6.1 Concept in Four Steps

The concept itself consists of four steps, where three of them directly impact the business. The first step is to analyze the information environment of the recommendation domains of the business in question. This is to be accomplished with the use of the Information Ecology Model (Rosenfeld, Morville & Arango, 2015) (all utilized methods will be reviewed in the *Methodology* chapter) in order to get a clear overview of their recommendable content and their relevant users, alongside their overall potential for possible recommendations. This will, for example, give insights into the type of environment that the recommendations have to be implemented into and the type of recommendation filter that the business might find most valuable for their specific situation. The use of The Information Ecology Model can also prove to be valuable to the company itself, if they are interested in getting an overview of the targeted information environment. When an understanding of the company's recommendation domain and theoretical recommendation system has been reached, the next step will commence. The second step will depend upon the company's existing usage of their data, and will look into how they collect, structure and track their existing data while also looking into the potential to acquire and track new data. This step is designed to help them with these aspects based on current best practices in the field, and of course tailored to their exact situation (which was investigated in The Information Ecology Model). When the first two steps have been completed, and there is a lot of relevant data to work with, it is time to start conveying the information properly, and to give the business new semantic insights. This is done in the third step with the help of a thorough formal ontology, which argues for logical relationships between data snippets from users and content. These relationships will give an understanding of what content to recommend to which user with a framework provided by the recommender system that has been chosen as the starting point for this data evaluation. This is done with a relevant ontology editor, which helps establish a domain that can be used as a future reference point to argue for the logical implication in any content recommendation in the targeted information environment.

This brings the solution to its fourth and final step, which investigates a persuasive perspective in terms of how to actually communicate the right content to the right user. If the formal ontology is a success, and a logical link between certain users and certain content can be ascertained, it is sensible to look into persuasion principles to figure out where, how, and when to actually communicate the piece of information. The point of this step is to investigate persuasive principles that can be directly linked to the ontology, so that the findings and insights, which are gathered from it with logical implications, can be strengthened by examining the communicative aspect of the recommendation possibilities, resulting in a proposal for more holistic guidance on how the business should recommend content.

6.2 Overview of the Steps:

1. Examine the targeted information environment for the relevant recommendations, with the use of the Information Ecology Model. Furthermore, this will determine an overall framework in terms of a fitting recommender system.
2. Make sure the business domain is properly collecting, structuring, and tracking data according to the chosen filtering approach in question.
3. Create a formal ontology in order to argue for logical relationships between data from users and content (which content would users likely interact with next, based on logical assumptions?)
4. Use these insights alongside persuasive principles to communicate the actual recommendations to users in a useful way.

6.3 Numerous Solutions to One Problem

The steps and principles of information architecture, which in combination become the concept, are carefully picked for the purpose of building a stronger and more logical way of recommending content. However, the practical value that the concept provides takes top priority. As such, the developed process and the principles they include can, in certain cases, be subject to change based on the needs and context of the company in question. Information architecture is a broad field and includes numerous methods and processes, which allow for multiple ways of solving a problem. However, the established concept, with its current steps and principles, is expected to provide the desired, valuable results as described. Establishing a static framework for the concept, in regard to its content, also allows for its findings to be comparable, enabling scientific knowledge to arise from the development and use of the concept.

6.4 Varying Practical Value of Results

It is important to be aware of the varying practical value that the concept brings to a company. Based on the company's outset, resources, context and prioritization, valuable practical findings of the concept may vary drastically, depending on the company. In one case, the concept may be able to establish a coherent way of utilizing and structuring data. In another case, it may only be able to provide miniscule valuable insights which differ from what the company already has established. With a new theoretical concept that attempts to fit into the practical world, this is only to be expected, as the implementation of the concept, and its use in practice over time, is exactly what will allow for the concept's strengths and weaknesses to become apparent. However, since the purpose of the concept is not only to design a recommender system, but also establish a semantic layer, the concept is expected to provide some form of relevant and valuable knowledge for practical purposes.

6.5 Prerequisites

For the concept to succeed, there are specific prerequisites which must be met:

- As the concept is based on principles of information architecture, the addition of information architects or ontology engineers as stakeholders to partake in the development of the concept is required.
- The companies in question are required to provide the people, who are responsible for the concept, with detailed information about their contexts, users and content in the information environment, which both parties will agree on in collaboration.
- The companies will also need to provide information on their current structuring, tracking and use of data in relevant specified areas, which both parties will agree on in collaboration.
- To ensure that the concept brings, and keeps bringing, value to the business in question, there must be sufficient investment into maintaining the ontology and algorithm of the recommender system. Ensuring well-managed maintenance will also assist the concept in its scalability, if relevant.

7 Methodology

This chapter examines the most relevant methodology associated with this thesis. These are explained in relation to their overall academic purpose and will be shown here in chronological order. There will also be arguments for why these methods were chosen over similar methodologies, and how they fit into the academic toolbox of the information architect. Finally, the limitations of the overall methodology will be accounted for.

7.1 The Information Ecology Model

The framework, which the analysis part of the concept is based on, is The Information Ecology Model (Rosenfeld, Morville & Arango, 2015). The Information Ecology Model is an analysis method often associated with the dissemination of information. It is therefore widely used in information sciences, and particularly in regard to data. The idea is to analyze three subdomains (*context*, *content* and *users*) of a larger domain, which also intersect with each other to create a holistic overview. This is a way to create a general understanding of the domain on the basis of the relationship between these subdomains (Figure 5). The decision to choose the Information Ecology Model comes from its flexibility and scalability for many different companies.

In this particular concept, the framework is also utilized in order to develop an understanding of the overall holistic domain for a given recommender system, based on the *context*, *content* and *users* associated with the company in question. Of course, this is only related to recommendation of content, since it is possible to collect and work with too much information, and thereby make the analysis less precise. With this knowledge from an in-depth analysis of the subdomains, it is possible to argue for certain design aspects of a recommender system for the domain in question, and eventually proceed to the next steps of the concept. Each of the subdomains provide an important perspective, which is primarily the *context* (market situation) of the business alongside the *content* and *users*, which are crucial to properly examine in relation to their possible role in a future recommendation service. The *content* and *users* will likely be examined in terms of groups in the domain due to structuring purposes of the recommendations.

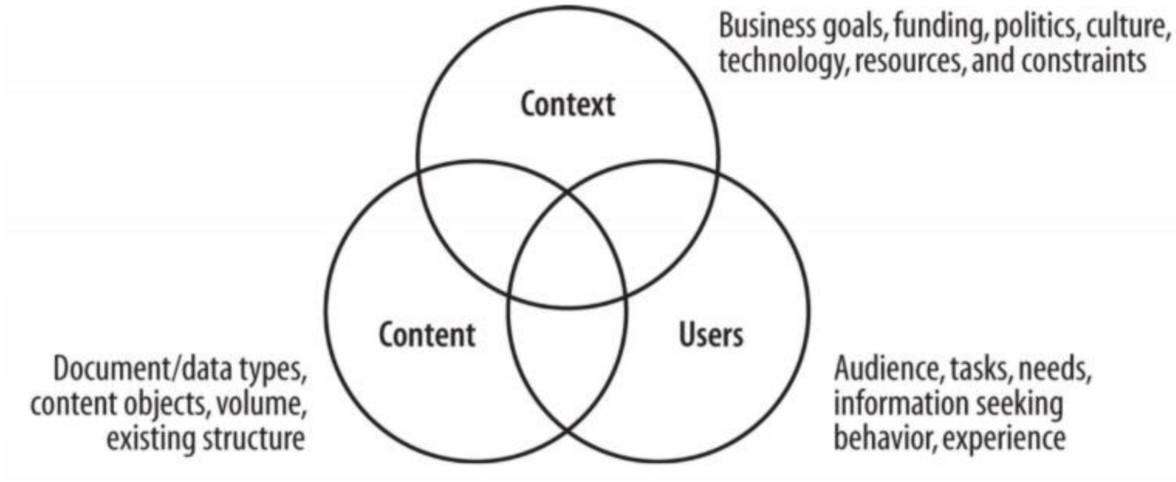

Figure 5: Information Ecology Model (Rosenfeld, Morville & Arango, 2015).

In the scenario of this concept, the subdomains that are being investigated are at the core of any of the possible recommendation filtering approaches. Specifically, the domains labeled content and user, which project the most important relationship factor in order to visualize possible recommendation examples. The holistic overview created by the context subdomain further strengthens the logical implications behind the user-to-content interaction. With proper usage of the model, it is much simpler to depict the exact recommender system that is most valuable to implement for a given business. Furthermore, the data analyzed can also create a basis for looking into how the company handles their data and how this all fits into a larger semantic architecture. With this in mind, the ecology model can be considered a strong and valuable tool for this exact thesis, since it is theoretically able to lay down a knowledge foundation for the entire recommendation service for basically any business regardless of size. The process for how each subdomain will be investigated is explained in the full description of the theoretical concept.

7.2 Ontology Methodology

In order to develop a working formal ontology, it is important to follow a set of development steps beforehand. Luckily, there is valuable methodological knowledge published regarding the development of formal ontologies. The methodology for this particular thesis is based on “Ontology Development 101: A Guide to Creating Your First Ontology” (Noy & McGuinness, 2001), which is a good starting point for developing an ontology to integrate into a semantic layer for a business structure. The methodology gives a plain and simple introduction into developing taxonomies and linking classes with different properties. It is an older methodology, but it still works today in that the process that it offers resonates well with digital implementation in an ontology editor.

There has been some critique in terms of applying the development process in small to medium-sized businesses (Öhgren & Sandkuhl, 2005). The critique mainly associates the development process with a lack of implementation possibilities, which are not included in the step-by-step guide. However, it is not a problem to include from other sources, and the specific one that has been chosen for this concept is no exception. The critique also mentions evaluation, which is also a weak point of this exact development process. In fact, it is mentioned as a general problem at the time that the article was written, however, it should be sufficient to evaluate value based on the answering of competency questions. Developing an ontology is also an iterative process, where steps are not always intuitively followed in a specific order. There might therefore be situations, where some steps are more likely than others to be implemented first, even though the sequence does not support this. The steps can therefore be considered a guideline.

The act of choosing this development methodology should not make the structure of the overall concept lack anything significant. With an established methodology, the development of a formal ontology can commence and afterwards enter the implementation phase with the use of the OWL language and Protégé. The exact steps that are used will be properly examined in the concept description.

7.3 Protégé

Protégé is an open-source ontology editor, and it is the tool that has been chosen in this thesis to implement and support the developed ontology ("Protégé", n.d.). OWL (Web Ontology Language) is the language that Protégé is based on. It is important that the ontology can be integrated into this language since it makes the workflow easier due to the simplicity of retracting useful information and arguing for semantic patterns across the selected ontological domain.

A particularly helpful tool regarding the use of the OWL and the Protégé tool is the possibility to use the reasoner-feature, which helps answering asked competency questions in collaboration with the development of the ontology. This is also useful regarding finding answers to new questions that may arise later on in the utilization phase regarding relationships between entities in the domain. Protégé is very flexible to edit with and has a simple learning curve. It has been reported that in Pinterest's case, it did not take very long for employees to grasp the core aspects of using Protégé to work with data (Gonçalves et al., 2019). This ultimately assisted them in saving substantial amounts of time and reducing the number of Excel spreadsheets that they needed to use to keep track of their domain.

Protégé is chosen for this thesis over other ontology editors, since it is currently popular around the globe ("Ontology editors", n.d.). This makes it likely that it is the ontology editor-tool that is most often encountered in other professional settings unrelated to this thesis. It is also regarded as pluggable and modifiable, which increases the amount of practically viable possibilities of the editor. Overall, Protégé and OWL can generally be considered a solid choice for this thesis and the concept featured herein, since it provides the perfect amount of flexibility and a holistic overview to help argue for the right content recommendation. It is also a safe choice that has proven to be fairly simple to get to grips with by people who are new to semantic technology. The utilization of Protégé will integrate nicely into the ontology development plan with the exact terms and capabilities, which were examined previously in the *Ontology Methodology* section (Horridge, 2011).

7.4 Behavior Model

In order to communicate the recommendations that are found and argued for with the ontology, it is relevant to utilize persuasion principles to secure a persuasive recommendation strategy in a practical setting. A highly viable method to utilize in this situation is the Behavior Model (Fogg, B.J., 2009), which gives an idea as to how to target the user-groups accordingly, so that there is a high chance of user-interaction with the content that is being recommended. It is important that the ontology has already been developed to a degree, where patterns related to recommendation potential between users and content are able to be recognized rather easily. From here, the question of how the pattern can be acted upon, and how practicality is being linked to the ontological realm, is important to investigate in-depth. Persuasion principles such as the Behavior Model are a great tool in this regard, since it gives very precise suggestions in a simple manner that are easy to implement into the overall algorithm.

The Behavior Model is particularly interesting compared to other persuasive tools, since it provides a very simple yet powerful way to get a holistic overview of the user and content relationship, and furthermore gives suggestions in regard to how the interaction might be strengthened and lead to a conversion. This is first and foremost done by examining the principles of *motivation*, *ability* and *prompts*, which are the key concepts of the model (Figure 6).

Motivation is on the Y-axis and constitutes the likelihood of interaction by the user based on how motivated they are. An example of a persuasive concept to create change on this axis could be *pleasure/pain* (a concept will give you pleasure/a concept will take away pain). *Ability* is on the X-axis and constitutes the ability to interact by the user. If the user does not feel like they possess the required ability to interact, then they very likely will not. A factor to alter here could be *time* (the ability-increase could happen from making the goal quicker, which in turn might make it more viable for the user). *Prompts* are the call-to-action interaction that is ultimately based on the collaboration between *motivation* and *ability*. In order for a *prompt* to be successful, it must be located in the upper-right corner of the graph, since this is where motivation and ability proves most likely to be interacted with. If the *prompt* is not located in a position, where there is a high likelihood of user interaction, the model then helps by suggesting that one might need to look into strengthening either *motivation* or *ability* to increase the success of the *prompt*.

There are some different types of prompts that establish themselves in the practical utilization of the model. *Spark* is a prompt that is accompanied by a motivational element.

Facilitator is a prompt created with the intention to also increase ability. Lastly, *signal* is a prompt that functions as a reminder for a sufficient user to be aware that they already possess the required motivation and ability. If the correct type of prompt is used, then the likelihood of user interaction is high.

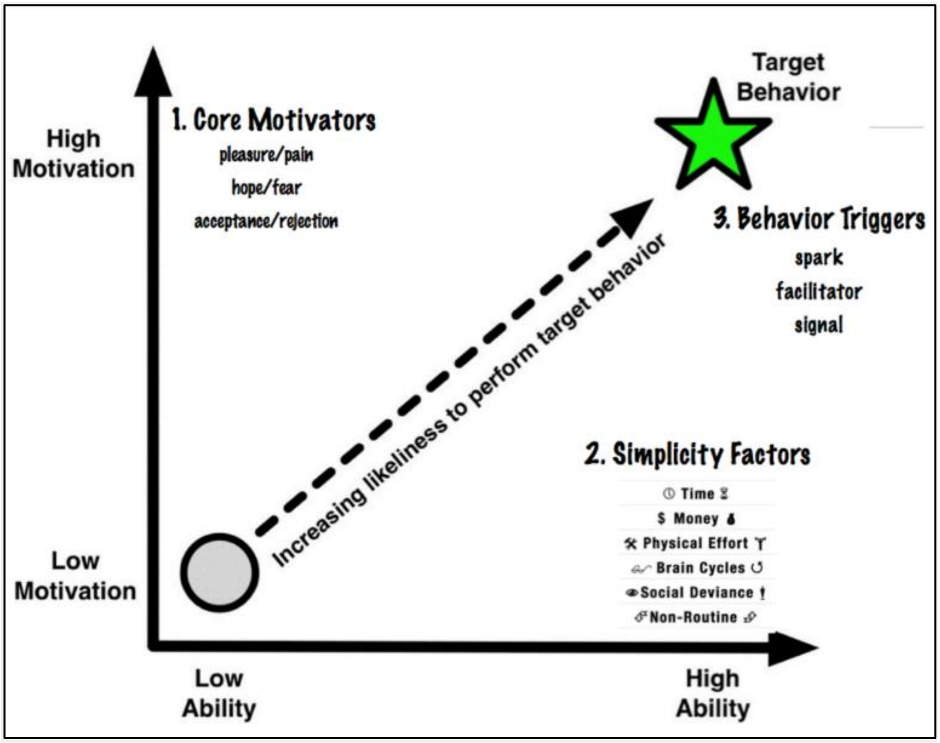

Figure 6: B.J. Fogg's Behavior Model (Fogg, B.J., 2009).

The way that the Behavior Model is supposed to work alongside the concept of the thesis is by using it to guide the actual communication of the recommendations. For example, take some of the prominent patterns between a particular user and a piece of content from the ontology. Then use the model to judge the amount of motivation and ability present in the interaction between said user and piece of content, and then further establish which prompt works best with the final recommendation example. If the factors are not sufficient for the piece of content to be recommended, it is an indicator to strengthen either *motivation* or *ability*. It is a very simple but powerful tool to utilize alongside the practical communication of recommendations.

7.5 Kairos

In most examples of providing digital recommendations in a practical setting, the recommendations will be implemented into a section of a website or an application that is deliberately made easy to spot while interacting with content. In other cases there is a possibility to provide recommendations that are much more in-depth in terms of communication, and will simultaneously become much riskier, but also much more rewarding for both the user and the business, if successful. A way to enhance the communicative aspect and specialize the recommendation is by utilizing time and place. This is also known as the Kairos-moment (Glud & Jespersen, 2008)

The idea is that an opportune moment is created by an objective and subjective interpretation of time and space. The *objective time* is astronomical while the *objective space* is based on coordinates, which cannot be altered. The *subjective time* and *subjective space* are, however, strange concepts and vary for each individual user. These are based on each individual user's consciousness, which makes it impossible to fully comprehend. A subjective interpretation of time would, for example, be a sense of understanding of what has taken place before, and likely will take place after, a given moment in time. It is something that is very hard to measure. A subjective interpretation of place is the sense of presence in a given place. If a person is physically present but has their mind fully preoccupied, or if they are distracted with something else, they might not be aware of their physical presence on a given set of coordinates. Together the objective and subjective interpretation of time and place creates the opportune moment, which can be acted upon persuasively (Figure 7).

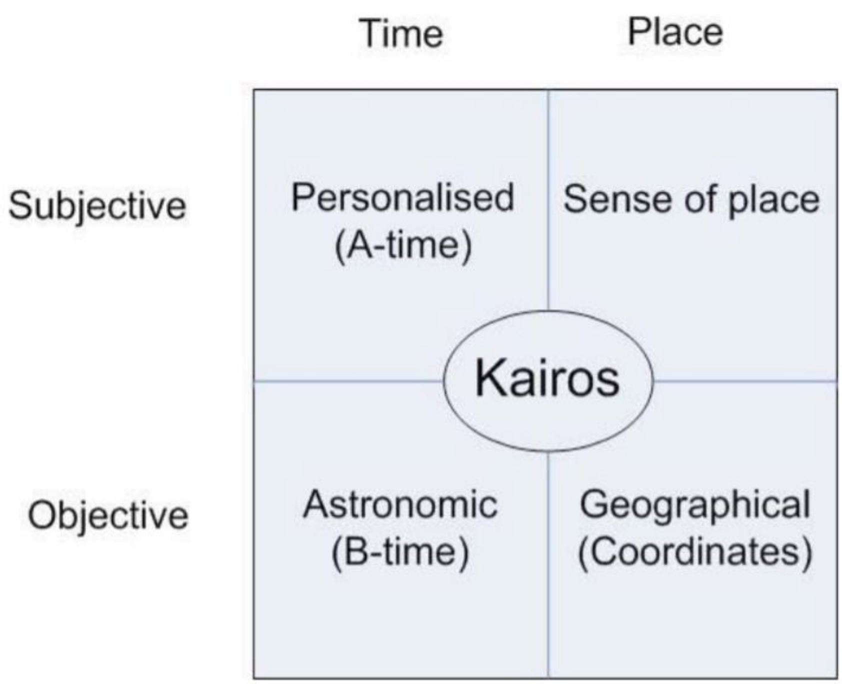

Figure 7 The definitive factors of the Kairos-moment (Glud & Jespersen, 2008).

In order to utilize the Kairos-moment with regard to persuasion on a digital domain, the four factors in the above model must be accounted for. A significant step is to obtain an understanding of a certain user's or user group's lifestyle and hobbies. This is done so that the persuasion can be based on the likelihood of a certain conscious placement of the user in *subjective time and space*. The *objective time and space* on the other hand, is much easier to get an understanding of, and can in many instances be tracked very precisely. If the user is using a mobile application, for instance, it is not a problem to collect this information alongside the data about the user interaction itself.

For the featured concept, Kairos can be utilized when accounting for push-notification recommendations that utilize a user's smartphone. There are also other possibilities, but this is one of the most common methods. There are also email recommendations and the like, but this is not a format that is possible to measure objective space on. As such, Kairos can be seen as a strong tool in persuasive technologies based on its risk/reward ratio, and is therefore very relevant for this thesis, where showing the practical aspect of persuasion is one of the main focal points.

7.6 Limitations

The methodologies that are used in this thesis are supposedly sufficient for providing a solution to the initial problem area and hypothesis. However, there is a lot of unused potential with these methods, based on the sheer number of different directions for the concept idea. In order for the concept to stay on track and create a valuable end-result on the basis of the methodology reviewed above, it has been important to appoint a set of limitations for the process to flow accordingly. In this section, the core limitations for the project will be examined.

Firstly, it is important to state that the approach to this thesis and concept is designing, and not creating (as in programming), a product. The idea at its core has never been to build a recommender system and actualize it throughout the thesis, but to examine the design of such a system and investigate in-depth the logic behind the information available for such a system. This is in line with what is expected from an Information Architect, since the scope is about conveying and working with information in ways that pave the way for the actual coding to be done.

Similarly, it is also important to state that the testing phase for the actual recommendations are difficult to handle in this thesis. The way the recommendations are rooted in the domain of the respective businesses makes it problematic to test before building the prototype of the feature. Because of this, prototype testing is not a part of the aspect of knowledge in this thesis. Instead, conclusions will be based on reviewing the practicality of the semantic layer's effect on a recommender system, and the persuasive aspect of the communication that this brings forth.

Limitations of the analysis concern the three domains of the Information Ecology Model that are being examined. These only relate to the recommendation possibilities of the given business that is targeted. It is possible to do a broader analysis that fits outside of these subdomains. This could be in regard to the market situation, benchmarking or comparing strengths/weaknesses, for example. The limitation here is set to make sure that the thesis and accompanying concept does not become too practice-oriented on the situation of a particular business and thereby losing the academic perspective that was chosen in order to investigate a problem.

Finally, in terms of the limitations of the concept, it is relevant to look at the featured persuasive principles to strengthen the recommendation; the Behavior Model and the Kairos-moment, respectively. These two principles fit into the recommendation domain, as they are able to help communicate the recommendations that are featured in the thesis and improve upon the end-result. There exist many persuasive principles, and the choice to ultimately go with these ones is rooted in their versatility. Other persuasive principles could be geared towards the user experience in general, such as gamification elements. Or trying to make the user invested in a product using the hook model for instance (Eyal, 2014). It is important to keep in mind that persuasion needs a purpose.

Only choosing two principles is ultimately a question of actively putting in limitations because the only domain of interest are the specific types of recommendations. Putting in more principles might even prove unnecessary to an extent; since the project needs to stay on scope, it is important that the persuasive principles featured are carefully selected for their specific jobs. The two principles are deemed worthwhile in this manner, since they fit the domain of recommendation well.

7.7 State of the Art

The concept that is featured in this thesis introduces, at the core of the theoretical course of action, the implementation of a recommender system in conjunction with a semantic layer, based on formal ontology. This is done in order to help argue for the logical implications associated with the recommendations. The overall purpose of the thesis is to shed light on the practical utilization of information architecture for businesses in general, however, this particular fusion of ontological structure and theories of recommendation engines associated with a digital domain is at the core of the research criteria to create actual academic value for the thesis. It is important to keep in mind that this is only considered a part of the overall argument for the value of information architecture, since the field is too broad to examine comprehensively in this thesis alone.

It is only through bite-sized research that an overall argument can be derived and furthermore processed to argue for the practical value perspective of information architecture.

The idea of implementing recommender systems into ontologies and vice versa is not an entirely new concept to be utilized academically (although not with the particular purpose that this thesis puts forth), and therefore it is important to establish state-of-the-art examples in order to create the best possible starting point for this particular concept and its use of some of the same principles. The plan is not to base the concept directly off of another state-of-the-art concept, but to get inspired for how to argue for its semantic value and structure the process accordingly. It is also important to use these examples to underline the practical value of this particular course of action, in order to further strengthen the argument of relevance in terms of utilization of information architecture for this concept.

Based on the different overall recommendation filtering approaches (which are presented in the *Recommender System* section of the concept description), it is relevant to present the *content-based*, *collaborative*, and *hybrid filtering* approach in conjunction with ontology development and implementation. In this regard, there are some interesting studies that have been done that directly combine the two in order to solve real world problems. They are not necessarily tied to any businesses but can be considered academic approaches that are rooted in researching a value potential.

One study investigates the general approach of recommendation filtering in an ontological perspective and deems it ultimately valuable in order to get new insights on problems regarding classic recommendation (Shah & Subramanian, 2019). “Classic” meaning the typical scenario of content-interaction on a webshop, for example. There is also evidence for search queries giving more useful results with this kind of implementation, which suggests that the ontological aspect helps to make advanced queries more comprehensible for the recommendation system to interpret (Thanapalasingam et al., 2018).

Regarding the filtering approaches, which are considered crucial for the content-interaction scenarios (and thus the concept of this thesis), previous studies have examined the *collaborative filtering* approach (the one that derives recommendation based on the existing user-base) in association with semantic data structure. One of the conclusions states that: “our approach not only outperforms traditional *collaborative filtering* in prediction accuracy but also offers improvements in coverage” (Sieg, Mobasher & Burke, 2010), which highly underlines the value-creating possibilities of this approach altogether. There is also data that suggests that the *collaborative filtering* approach can more easily overcome one of its largest obstacles in collaboration with formal ontology, which is the infamous cold-start problem (Sheridan, Onsjö, Becerra, Jimenez & Dueñas, 2019).

Hybrid filtering in a recommender system showed better results and higher recommendation satisfaction based on evaluation surveys (Ibrahim et al., 2018), while an investigation of *content-based filtering* showed that in an ontological domain that was properly constructed (with the use of OWL etc.), there was a rich possibility to argue for new snippets of knowledge, which were called “hidden semantic associations” between users’ preferences and content available (Lops et al., 2011). Overall semantic web technology seems worthwhile to implement in collaboration with the most popular recommendation filtering approaches that exist on the market today, in order to strengthen the recommendation quality that gets delivered to users.

The examined papers have been investigating the different filtering approaches that are relevant for this concept, and there are no reports of semantic structure having no valuable benefits for the final recommendation potential in their specific cases. This makes it likely that the logical approach is generally worthwhile, and if it turns out that it is not, it is an insight in its own right. Therefore, it can be deemed a worthwhile subject to research, to further strengthen this approach professionally, and make it more practically useful as well.

8 Concept Description

This chapter will expand upon the initial concept, which was introduced briefly in the *Concept Introduction* chapter, and further examine the process theoretically in a way that will subsequently be easy to apply in a practical setting. In this chapter, different steps will be fleshed out and investigated thoroughly, so that there is no doubt as to how an actual business could be assisted in applying this particular concept.

Recommender systems exist to reduce information overload and to help give users a better overview of their interaction possibilities on a given domain. This is why they were introduced into digital platforms in the first place. As one of the purposes of a recommender system is to estimate ratings for items that have not yet been rated by a user (Smetsers, 2013), the purpose of the concept is to methodically provide the design for a recommender system that is the most effective and useful for the particular company's information environment. Before the first step of the concept can commence, the domain, in which the concept is to operate, must be defined. The information environment for the concept is decided upon by discussing the purpose of the recommender system as well as the utility of the system, narrowing down the scale of the domain. Defining the information environment is also accomplished by weighing the company's wishes, needs, and resources against the consultant's knowledge of the concepts' capabilities. With a defined information environment, the recommender system can be designed. This leads us to the initiation of the first step of the concept, but first a holistic presentation of the different steps of the concept will be examined to create an overview of the overall work process and relationship between the client and consultant.

8.1 Concept Model

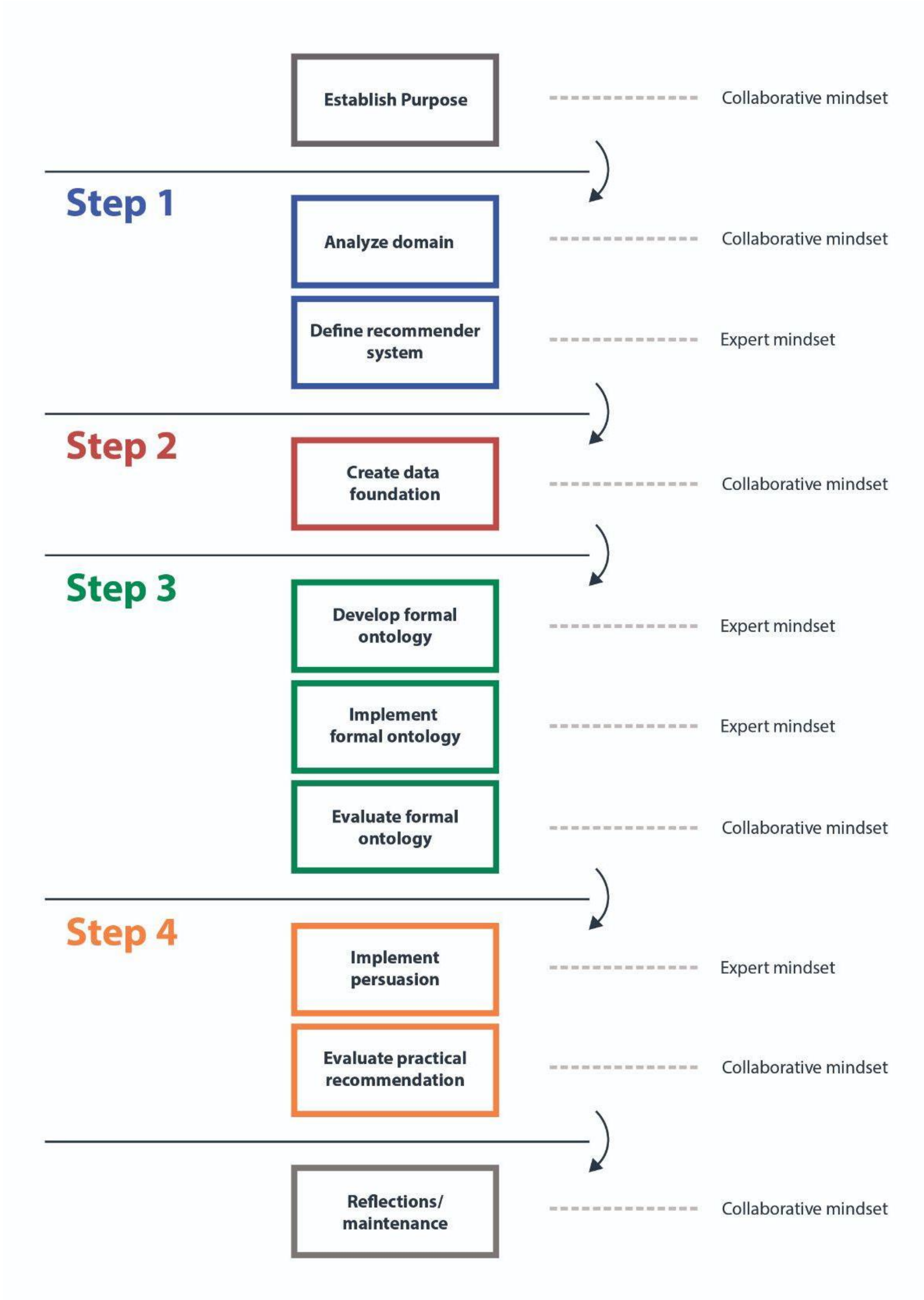

Figure 8: The concept description depicted in steps.

This model (Figure 8) shows the walkthrough depicted in individual steps with substeps to make an in-depth visualization of the sequence of actions. The work process of the concept does not start with *Step 1*. In practice, it starts beforehand, with the initial meeting between the consultant and the client. This is where the purpose of the recommender system is going to be discussed. This also includes the initial idea as to where the recommendation is going to be placed. For example, with a push notification in an application, or with a content presentation window on a website. After a sufficient amount of knowledge has been shared, the consultants can move on to *Step 1*. This step acts as an analysis, and an initial definition of the potential recommender systems' capabilities, based on the information from the analysis.

Note the inclusion of different *mindsets* on the right-hand side of the model. This depicts whether the people responsible for the defined domain are a direct part of the step. In other words, if the business and the consultants need to work together in order to establish the best possible outcome. If this is not possible, the outcome of the step may end up not meeting the wishes or requirements of the business. With a *collaborative mindset* being where cooperation is required to proceed, an *expert mindset* indicates that the step can be evaluated and worked through by the consultants alone (with the condition of having previously obtained sufficient knowledge).

Step 2 is a careful look into the information/data possibilities for the company. Are the optimal data handling principles for the recommender system already in place? Or do new practices need to be established? *Step 3* delves into the semantic structuring and usage of the knowledge that has been gained so far in the concept. It accomplishes this by dividing the engineering of an ontology into three different substeps: The first substep is the development and establishment of the defining elements of the ontology such as classes, properties and so on. The following substep is the implementation of the ontology in an editor. The final substep is an evaluation of the ontologies' reasoning capabilities by answering the initial competency questions. S*tep 4* introduces persuasion to the concept. In order to get the best idea of how to implement persuasion to finalize the practical recommendations, this step has to be completed after the semantic layer has been established. *Step 4* ends with an evaluation of how the actual recommendation is being communicated. This also summarizes the overall product that the concept is able to define. Lastly, there is room to reflect upon insights from the whole process and decide on maintenance possibilities for the company. With an overarching plan of the concept, the specific content of the steps will be examined in-depth.

8.2 Step 1 - Business Analysis

This is a fundamental part of the concept process, where thorough analysis will help create an understanding of the chosen domain and establish what kind of recommender system will fit the business' needs most optimally in conjunction with a semantic layer.

8.2.1 Information Ecology Model

The very first thing that must be done in order to get a proper understanding of a business, which has chosen to seek guidance regarding optimization of recommendation capabilities, is to investigate the surrounding subdomains of the chosen domain where the concept is expected to operate. *Context*, *content* and *users* are the names of these subdomains. This investigation is focused specifically with regard to gaining an in-depth understanding of the recommendation possibilities at hand. This entails a process, whereby *context* will investigate the specific context in which the business needs recommendation and the maintenance possibilities associated with this aspect. Then *content* will investigate the content that is present in the domain, and its potential to be recommended to users. Finally, *users* examine the different user groups of the domain regarding the reception of recommendations based on a number of different criteria.

8.2.1.1 Context

The typical use of this subdomain is to highlight the overall context of a given business concerning their position in the market, and their strengths/weaknesses compared to similar businesses etc. In this case, however, it is important to investigate the context of the impact that the recommendation has on the business. In other words, what is the context between the business (their data or potential to acquire data) and the conceptual meaning of "*to recommend content*" in their case? How is it relevant for them? How does it create value? What measures must be taken in order to undertake the assignment of creating and overseeing a recommender system etc. There is a wide range of subjects to delve into when it comes to the context, since recommendation can be done in many different ways, which have the potential to create a certain amount of value. The more questions that can be answered, the better the opportunity to understand and guide the client towards the most sensible and reasonable angle of recommendation, ensuring that the approach aligns with their business models, strategies, and values. This is important for the purpose of creating practical value. It is also important to look into the context of utilization and maintenance of a finished semantic recommender system through the use of an ontology in the business domain, to ensure future scalability of the system.

8.2.1.2 Content

This section of the Information Ecology Model is usually rooted in examining existing content and how it is interactable on the given digital platform that it resides on. This is also relevant for this concept. However, it is just as important to examine the content in a way that goes in-depth with how the different content groups function in a recommendation setting. What type of content does the business provide in the domain? How is the different content presented to the user? How is the content grouped into categories and how do these groups impact each other? Answering these questions are relevant when it comes to understanding how the content fits into a larger understanding of recommendation. However, simply understanding the domain's content and the purposes for it, is not enough to gain the necessary understanding of the recommendation domain. It is also important to understand the relationship between content and the corresponding users, who ultimately benefit from the possible recommendations.

8.2.1.3 Users

This section examines the subdomain of *users*, including what exactly defines the users of the established domain, and how these particular users tie into the available recommendation capabilities. The users can be examined from many different perspectives, but for the purpose of this thesis, it is important to align the user analysis with the recommendation capabilities. What are the user groups that are present in the company's domain? How do they interact with each other? What can be utilized from the data from each individual user and their previous interactions with the content? This, alongside the aforementioned investigation into context and content, paves the way for an in-depth understanding as to how to approach recommendations in the established domain.

With all three perspectives it is possible to look at the grand scheme of the chosen domain and examine practical approaches to recommendations. What might be the right choice to benefit one company, may not be the same for another. As such, each subdomain must be carefully considered. With an understanding of the domain through the Information Ecology Model, an initial definition of the recommender systems capabilities can be initiated.

8.2.2 Recommender System

With an established domain, and a knowledgeable perspective of the *context*, *content* and *users* through the use of the framework of the Information Ecology Model, the appropriate capabilities of the recommender system can be defined. One of the most important parts of defining the recommender system is to decide on the filtering approach. However, before a filtering approach can be chosen, the surrounding contexts regarding the nature of the recommender system itself must be made.

The data that the recommendation engine is supposed to look at and collect must be discussed and defined. Recommendation engines usually collect data from customers' search queries, their purchase history, social behavior, geographical location or demographics ("5 Best Practices for Effective Personalized Product Recommendations", n.d.). Based on the knowledge from the analysis of the domain with the Information Ecology Model, it becomes possible to define the type of data that the recommendation engine is supposed to collect. Knowing this will support the work process of the concept when examining the data handling principles in *Step 2*.

With the knowledge gained from the Information Ecology Model, the device-platform, from where the system is to be based, can also be determined. For example, in the case of a domain that focuses solely on mobile devices, mobile recommender system practices may be effectively introduced to the system. This is because rules and practices regarding data collection on mobile applications can differ from the rules regarding data collection on computers (Ricci, 2011). With these surrounding contexts of the recommender system, the selection of the appropriate filtering approach can commence.

8.2.2.1 Different Filtering Methods

Deciding on a proven filtering approach assists in ensuring that the recommendations that are communicated are valuable for both the users and the company. Choosing the appropriate filtering approach should be based on detailed information about the domain, hence why the analysis from the Information Ecology Model is needed beforehand.

When deciding on the filtering approach, specific variables of the recommender system like its accuracy, efficiency, stability, transparency, and serendipity must be considered based on the formerly established knowledge of the system. This is because the different methods for filtering vary in their utilization of each variable. For example, if transparency is an important variable to the context of the recommender system, an item-to-item approach might be better suited than a user-to-user approach. Informing your user that they have received a recommendation for an item, based on a list of similar items that the user has engaged with prior to the one that is recommended, it is simple for the user to understand why the recommendation is happening. When informing your user that they have received a recommendation for an item, based on another user's choices, it is difficult for the user to decipher. This is because they do not know about, nor have a relationship with, the other users of the domain, from which the recommendation is based (Smetsers, 2013).

Being able to choose the appropriate filtering approach on the basis of the gained information from the analysis, requires a strong knowledge base of the available types of filtering. In the case of this thesis, Narem Katakam's article on "How Can We Design an Intelligent Recommendation Engine" (Katakam, 2019) gives a general and cohesive overview, as well as visual representations of each filtering approach. Therefore, Katakam's visual models will be utilized in the explanation of the different filtering approaches. However, much of the overall knowledge base for this thesis' understanding of the filtering approaches is broader than Katakam's explanations. As such, the knowledge base also consists of:

- "Data-Science Recommendation System using Semantic Technology" (Shah & Subramanian, 2019) and "Ontology-based recommender system in higher education" (Obeid, C., Lahoud, I., Khoury, H., & Champin, P. 2018) for gaining a general understanding of common filtering approaches and their workings.
- "Content-based Recommender Systems: State of the Art and Trends'' (Lops, P., de Gemmis, M., & Semeraro, G., 2010) and "Trends in content-based recommendation" (Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. 2019) for gaining specific insights into content-based approaches.
- "Recommendation Systems Based on Association Rule Mining for a Target Object by Evolutionary Algorithms'' (Varzaneh, H. H., Neysiani, B. S., Ziafat, H., & Soltani, N. 2018) and "Ontology-Based Collaborative Recommendation" (Sieg, Mobasher & Burke, 2010) for gaining specific insights into collaborative approaches.

With this knowledge in mind, the different filtering approaches can be discussed through the visualizations of Katakam.

8.2.2.1.1 Popularity Filtering

The *popularity filtering* model is very straightforward. The content with the most views, likes, or conversions should be recommended to all users. As such, this way of filtering ensures that the most popular (and to some degree, successful) content is pushed on the users, resulting in the content's popularity rising even more (Katakam, 2019). An example of this would be a website for watching TV series that only recommends their most popular and successful shows to all their users. The problem with this model of filtering is that the recommendations never become personalized to the users, as all users receive the same recommendations, no matter the user's context or personal interests.

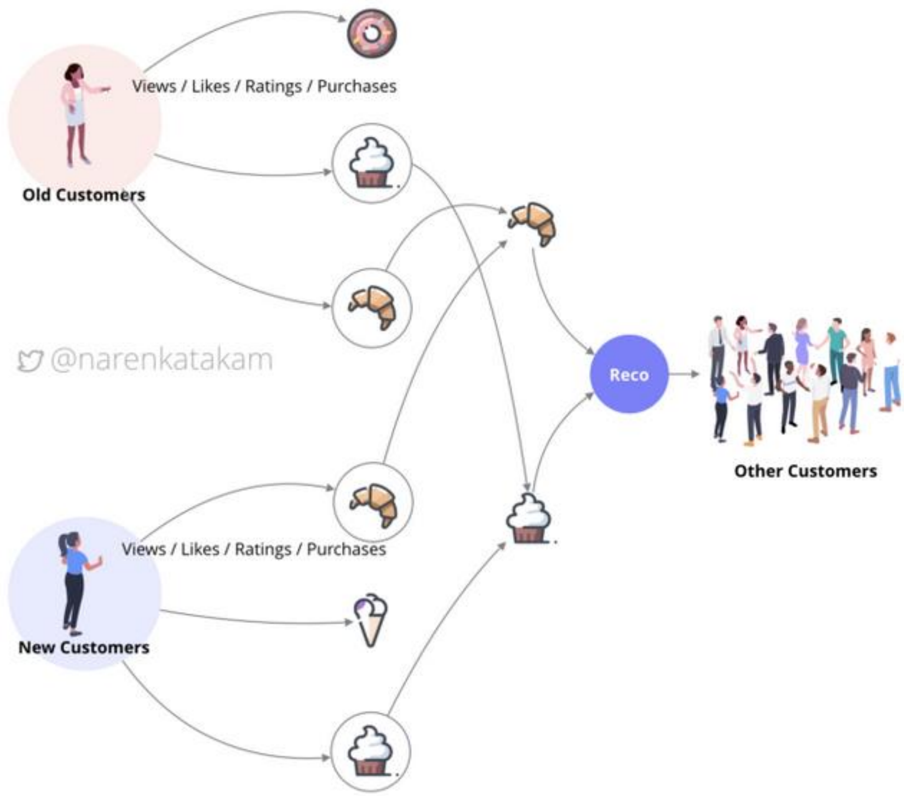

Figure 9: Visualization of the popularity filtering model (Katakam, 2019).

As is seen in Katakam's model (Figure 9), the figure illustrates how two user groups have existing relationships with different kinds of items. The two user groups share some of the items they *view*, *like*, *rate*, or *purchase* (they interact with the same type of item). Therefore, the system deems these items to be the most popular ones and recommends them to other customers.

Katakam's model manages to explain *popularity filtering* in a basic way which can easily be translated to this thesis. If the topic were about specific items in an IT-domain, rather than a domain of confectionery, the structuring of the model would remain the same.

8.2.2.1.2 Collaborative Filtering

The *collaborative filtering* model (Katakam, 2019) starts to personalize recommendations by basing the recommendation flow on different user profiles, which are established by collecting data about the users themselves. The model is based on the logic that people who agreed in the past, will agree in the future. An example of this would be how Netflix will provide a user with recommendations for different movies, based on similar users' choices. The presentation will often be seen with descriptive text, such as: "other users enjoyed these movies''. This model allows for personalized recommendations while still being able to recommend specific content to a large audience who share the same values. To accomplish this, the *collaborative filtering* model needs a lot of information about its users early on in order to be able to succeed and produce relevant recommendations. As such, the model is prone to run into the cold-start problem (Sheridan, Onsjö, Becerra, Jimenez & Dueñas, 2019). The limitations of the model become clearer as the performance of the recommender system becomes slower with an increasing user base. With the increasing user base, a lot of computational power is needed to encompass all the collected data.

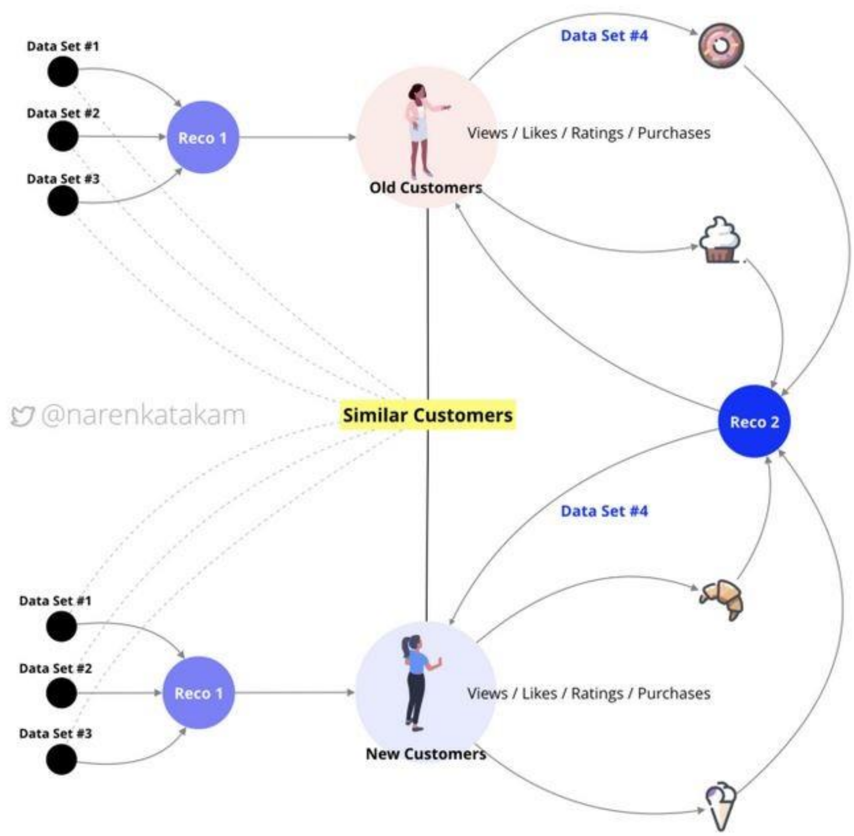

Figure 10: Visualization of the collaborative (Katakam, 2019).

Katakam's model (Figure 10) visualizes how two different user groups (or users) have datasets. These datasets can consist of almost anything about the user, such as demographic, geography, preferences etc. In the model, it is seen that both user groups share similar datasets. This makes the algorithm deem that even though the two user groups show interest in different items, they can be recommended items that the other user group has shown interest in. The reason for these recommendations is because the algorithm focuses more on the users' similarities, than the items' similarities.

Katakam's model succeeds in giving a general overview of *collaborative filtering*. However, the visualization would be easier to understand if the element *Reco 1* was removed and the datasets pointed directly to the user groups. The datasets directly represent information about the user groups, and therefore, when the user groups have similar information about them, patterns emerge, and from those patterns, recommendations can be created. When explaining *collaborative filtering*, there is no reason for the datasets to point towards a recommendation before pointing to the user group.

8.2.2.1.3 Content-based Filtering

Unlike the *collaborative filtering* model, the *content-based filtering* model focuses on the items (content) and not the users (Katakam, 2019). Typically, with *content-based filtering*, a strong similarity framework is built between the different types of content via metadata tags. When a user engages with a piece of content that contains specific metadata tags, the algorithm is able to recommend other kinds of content to the user, as long as the recommended content contains some of the same metadata tags as the originally engaged piece of content. Therefore, the algorithm builds content profiles with keywords and constructs user profiles that fit the content based on these tags. This means that it is not similar users' behavior that defines the recommendation, but it is the specific user's own actions that create the user profile, and as such, indirectly determines what kind of content is to be recommended to them.

The relationships between content are often more stable than relationships between users. As a result, the filtering does not become more complicated nor does it demand that much more computational power with a growing user base. Utilizing Netflix as the example once again, the *content-based filtering* model is seen when a collection of horror movies is recommended to the user based on the user's earlier interaction with another horror movie. This type of filtering model also makes it easy to explain to a user why they have received the specific recommendations in the first place. In the case of Netflix, a recommended list of movies will often be followed by a descriptive text like: "Because you watched *A Nightmare on Elm Street*". This explains clearly to the user why the recommendation is occurring. However, the *content-based filtering* model is limited by not being able to consider the quality of the content, only its similarities. This results in *content-based filtering* often being used in hybrid systems, rather than in isolation.

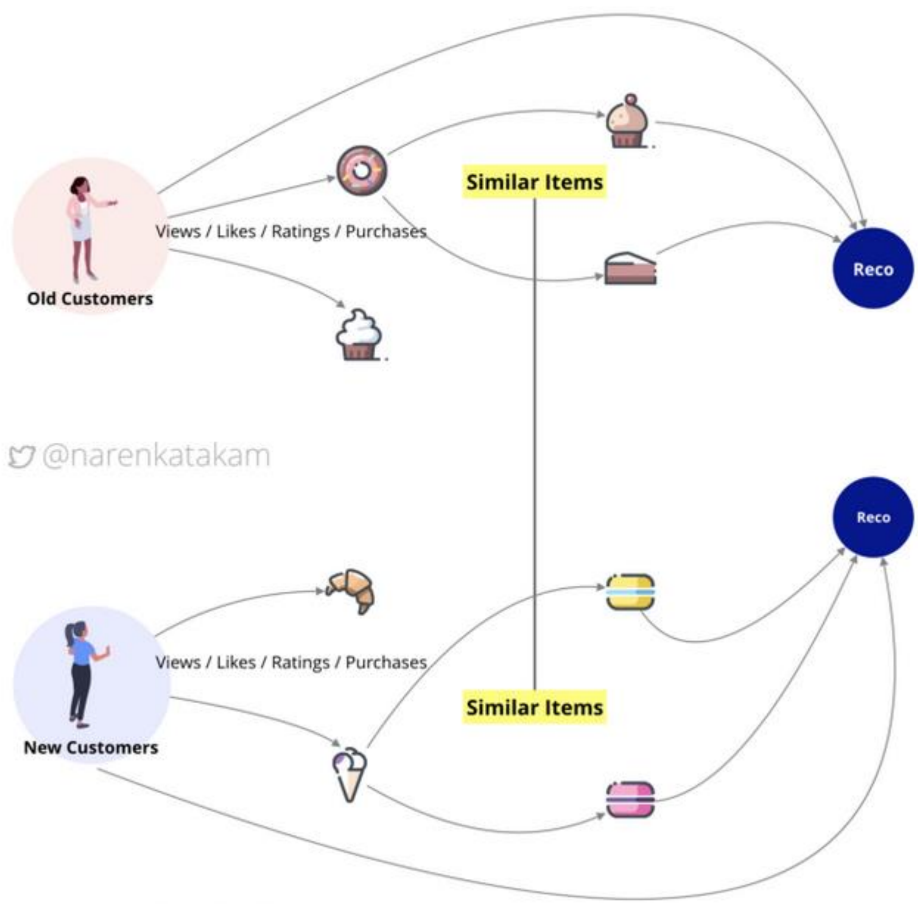

Figure 11: Visualization of the content-based filtering model (Katakam, 2019).

Katakam's earlier visualizations of the filtering approaches have given clear overviews of their specific approach, in line with the knowledge that has been presented in the article and broader knowledge base of the thesis, however Katakam's visualization for *content-based* approaches (Figure 11) could be misinterpreted, as the visualization is lacking in its illustration of the filtering approach. The visualization is meant to be understood as having two user groups with interests in different items. These items are related to other items by similarity (through similar item-profiles). As such, a user group can be recommended an item they have not yet shown interest in because an item they engaged with earlier has similarities to the item that is being recommended. This description of the model is purely speculative. The problem with Katakam's visualization of this filtering approach is that not a lot of contextual information is given. The same can be said for the two other models that have been established. However, they did not require contextual information in the same way that the *content-based* model did. To showcase this, a reconstructed visualization of Katakam's model has been designed (Figure 12).

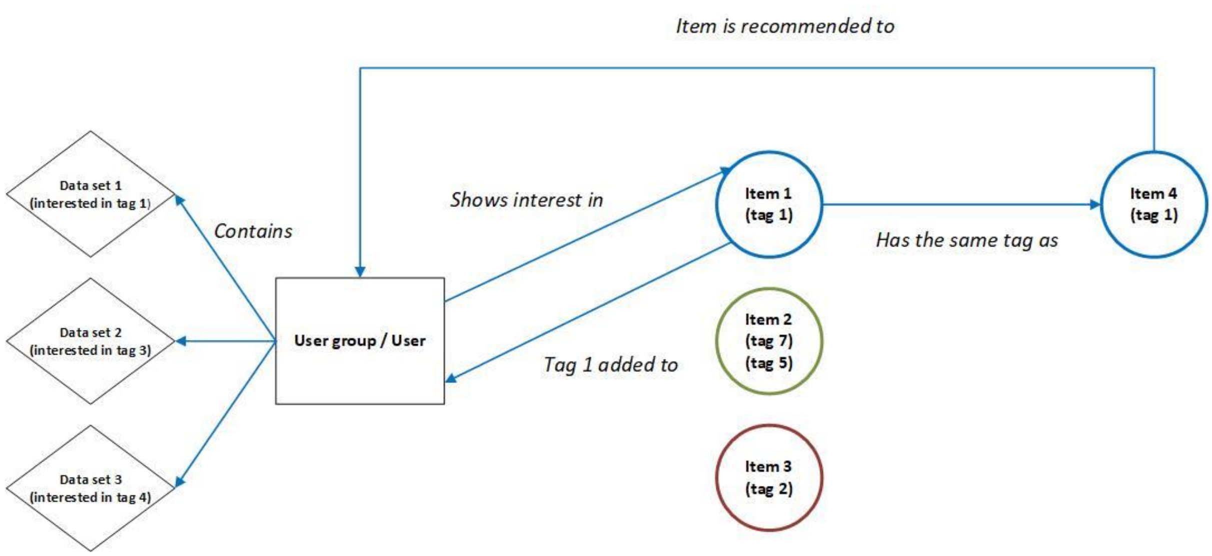

Figure 12: Reconstructed visualization of the content-based filtering model.

This new reconstructed model fixes some of the problems that exist within Katakam's model. Firstly, for the purpose of showcasing *content-based* recommendation, there is no need for two user groups. Therefore, the new model contains only one. Besides that, the new model contains a lot of contextual information. Each item now contains a certain metadata tag. The user group shows interest in a specific item, and this item contains a metadata tag which another item also contains. Therefore, an item that the user group has not shown interest in can be recommended to them. This is because the item they originally showed interest in contains the same tag as the recommended item. By showing interest in an item with a specific tag, that tag is applied to the user group's profile, containing different datasets about interests. There are other ways in which *content-based filtering* can occur (without metadata tags). However, showcasing the filtering approach this way, gives a clear representation that is easy to understand.

8.2.2.1.4 Hybrid Filtering Model

The *hybrid filtering* model does not contain specific rulesets for filtering like the other models. However, it combines different models to create a custom filtering model that encompasses the needs of the algorithm (Katakam, 2019). In many real-world scenarios, the *hybrid filtering* model is the obvious choice, as the combination of different models often results in stronger and more personalized recommendations. Most modern recommender systems utilize a *hybrid filtering* approach, as is the case with Netflix. As the *hybrid filtering* model combines multiple filtering models, it does not have the same limitations, as the intertwined models can assist in solving each other's limitations, which is the reason why, in many cases, the *hybrid filtering* model will seem like the obvious approach. One limitation of the *hybrid filtering* model is its complexity. The model shines in large, complicated systems but can be too ambitious and unnecessary in smaller, more basic systems.

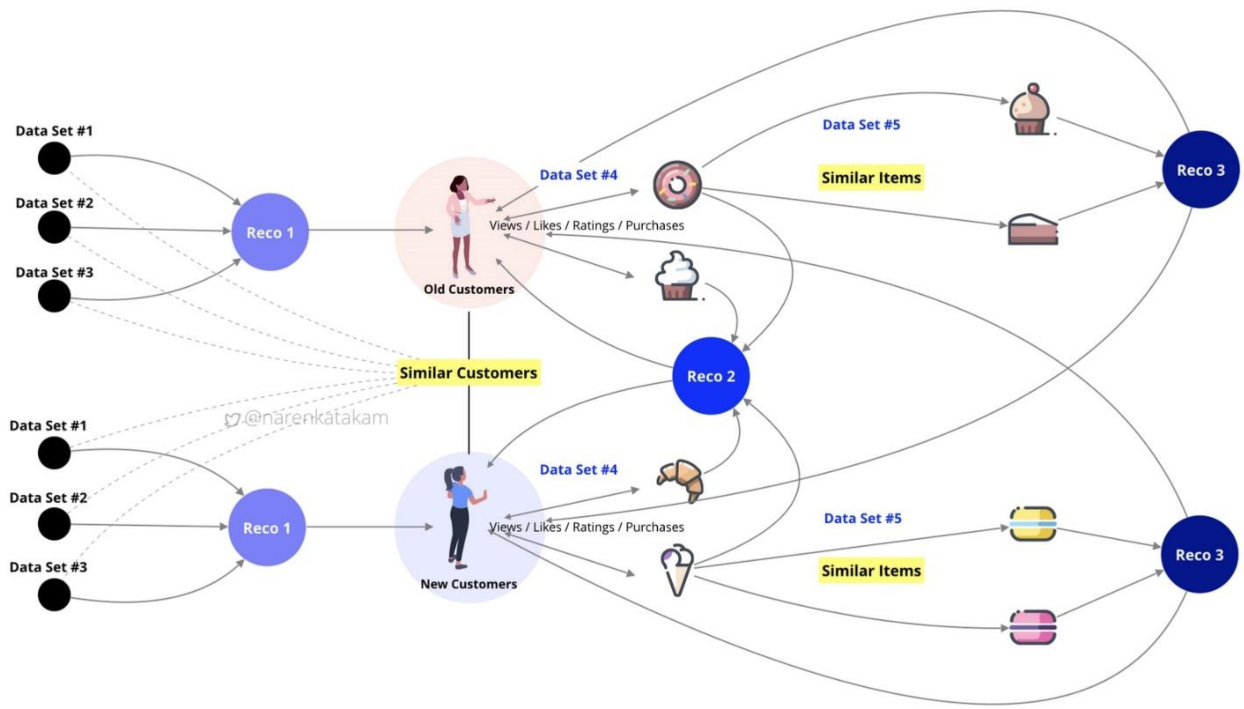

Figure 13: Visualization of the hybrid filtering model (Katakam, 2019).

Katakam's visualization of *hybrid filtering* (Figure 13) combines the *collaborative filtering* and *content-based filtering* visualizations into one. Even though the *content-based* visualization had to be reconstructed in the *Content-based Filtering* section, the *hybrid filtering* model's visualization of the *content-based* approach works, as its recommendations link back to the defined user groups. It still lacks the contextual information, and as such, it is not clear why the items are related. However, with this iteration, the recommendation (*Reco 3*) is pointing back towards the user groups, closing the loop, and making the model understandable. Regarding the rest of the *hybrid filtering* model's visualization, recommendation elements (*Reco 1*, *2*, and *3*) should all be understood as possible recommendations to be made. Katakam states that the assumption regarding *hybrid filtering* is:

"The proposed 'best' recommendation to a user is the one that other users with a comparable profile in a comparable state who chose the same or a similar product." – Katakam, 2019.

Following this statement, it is clear that the *hybrid filtering* approach has the most opportunities to provide personalized recommendations. Depending on the level of personalized recommendations needed, the best chance of providing the most optimal recommendation for a given user may be *hybrid filtering*. However, *hybrid filtering* demands a lot of information be weighed against each other. Therefore, according to Katakam, machine learning is a useful tool for ensuring that the algorithm provides optimal recommendations. Katakam has introduced this aspect in his visualization of the *hybrid filtering* model as well.

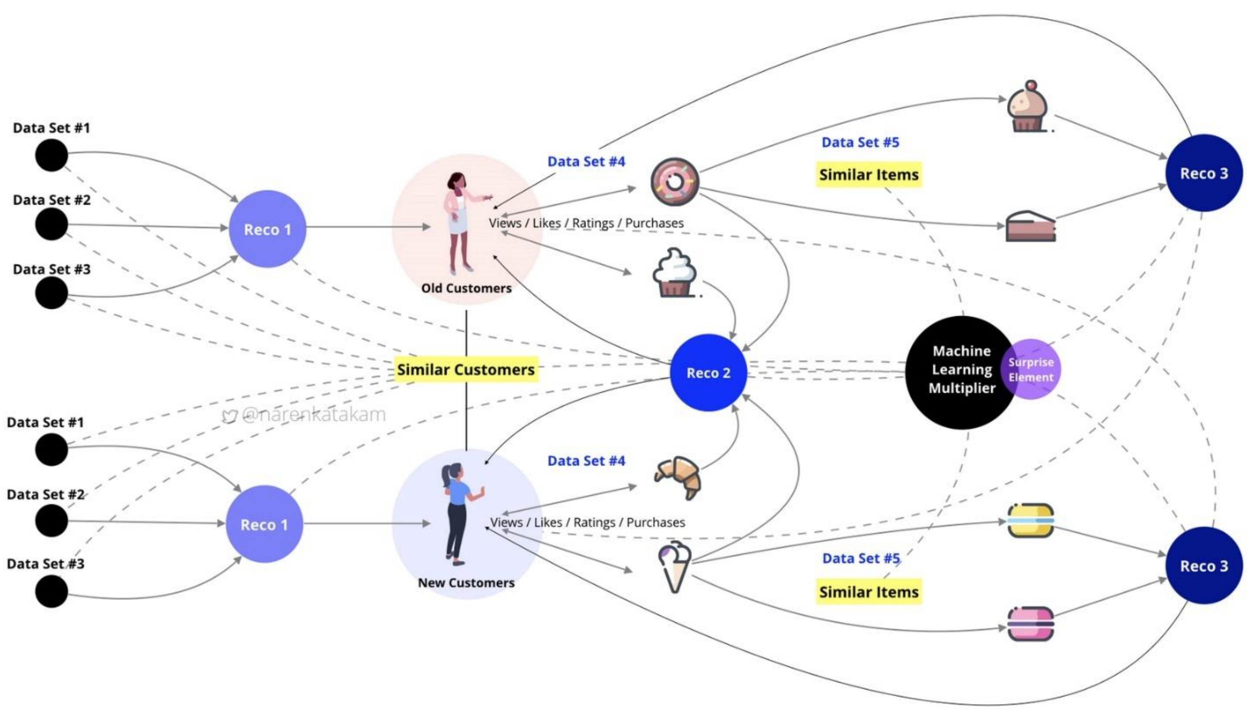

Figure 14: Visualization of the hybrid filtering model with machine learning (Katakam, 2019).

Even though Katakam introduces automated machine learning (Figure 14) as the tool for ensuring that the algorithm provides the most optimal recommendations, the consultants will utilize formal ontologies instead as an alternative. Formal ontologies will act as a “top-down” approach to recommendations, ensuring that logical and clear explanations for recommendations are present. While machine learning approaches can be very effective in this scenario, they are also ambiguous and act as a “bottom-up” approach, unable to provide clear and logical explanations for their recommendations.

When deciding on the approach for filtering, it is important to consider that even though the *hybrid filtering* model solves some of the limitations that the *popularity*, *collaborative*, and *content-based* model has by combining them, the other models have ways of dealing with their own challenges as well. As an example, with *collaborative filtering* approaches, some of the limitations of the approach can be overcome by introducing specific methods:

- Traditional Method based on neighborhood
- The Clustering Method
- Case-based Reasoning
- Compound Algorithms
- Association Rule Mining (ARM)

As the *collaborative* model becomes customizable in its attempts to solve its challenges, it makes the case for how *collaborative* approaches might be better suited than *hybrid* approaches in some cases. The same assumption can be made for the other filtering approaches, if applicable. In the end it is the purpose of the recommender system, and the filtering approach's capability to encompass the surrounding contexts of the recommender system that defines which filtering approach to choose (Varzaneh et al., 2018).

8.2.2.1.5 Additional Filtering Approaches

Besides the four aforementioned approaches for filtering, there also exists filtering approaches for other purposes and situations.

- *Session-based approaches* focus on ongoing sessions of users and adapt recommendations based on a user's actions in the specific session. Typically, the approach attempts to predict a user's next action in the session, or to point the user towards a predetermined action (Ludewig & Jannach, 2018).
- *Knowledge-based approaches* are based on a large knowledge record of users, content, specific content, and the needs of users (Varzaneh et al., 2018).

- *Memory-based approaches* are based on a database of users' known preferences for each piece of content and each prediction (Varzaneh et al., 2018).
- *Demographic-based approaches* base their recommendations on the demographic niche of a user profile (Obeid et al., 2018).
- *Community-based approaches* create recommendations for a user based on other closely connected users' preferences (Obeid et al., 2018).

While approaches like these are often more niche in their use than *popularity*, *collaborative*, *content-based*, and *hybrid* approaches, they are still highly effective for their purposes. Even though other approaches to filtering exist, the aforementioned approaches are more commonly used and still allow the concept to introduce the most optimal filtering approaches in a broad range of solutions.

8.2.2.2 Circumstances of the Recommendations

When the approach for filtering has been chosen, different circumstances of the recommendations must be decided upon, like with the number of recommendations to present for a user. In some cases, the best solution for the system will be to only recommend one specific item for a user. However, it is often the most beneficial option to recommend multiple items simultaneously. This can be done by recommending a planned sequence of items to a user, or by recommending the item that is the appropriate next step of a sequential process or session that the user is already engaged with. The recommendations can also be presented in a bundle, like when recommending a travel plan with flights, hotels, places to visit etc. (Smetsers, 2013). The circumstances of the recommendations involve a risk to the user experience, as recommendations themselves can result in an intrusive experience for the user. As such, many modern recommender systems are *risk-aware* or *context-aware* (Bouneffouf, Bouzeghoub, Gançarski, 2013), meaning that besides recommending content, they also consider the circumstances of the recommendation. Where should the recommendation be presented? When should it be presented? How often should it be presented? These are all important choices that will have drastic repercussions for the recommender system. How to recommend a piece of content to a user may differ based on the company's (domain's) needs, as well as the type of user, the type of content, and the problem that the system attempts to solve.

From the perspective of information architecture, making these choices for the recommender system is about making the system persuasive to the user. The Kairos-moment (Glud, Jespersen, 2008), regarding the opportune time and place, is an example of how persuasive design is able to assist or take over some of the aspects that makes a system risk-aware. In terms of the sequential concept process, an estimate on the circumstances of recommendations will be made before the next step of the concept will commence. The way this is executed with the concept is that the risk-awareness of the system is defined in *Step 2*. When *Step 4* commences, the possibilities of introducing persuasion into the case is determined and will support some of the responsibilities of the risk-awareness. In the later development of the concept, persuasion principles will be investigated to assist in reflecting and optimizing the circumstances of recommendations. With the surrounding contexts, the filtering approach, and the circumstances of recommendations having been defined for the recommender system, the next steps of the concept can proceed.

8.3 Step 2 - Proper Data Handling

After the information environment of the business has been thoroughly investigated, and a plan has started to form with regard to how potential recommendations could optimally be made, it is crucial that the client in question is properly aligned with their data capabilities. With a determined filtering approach that fits the domain and purpose of the recommender system, it is important that the data which the company possesses can actually be utilized by the chosen algorithm. Therefore, it is important to make sure that the business is collecting data about their users properly and structuring the data according to best practices and tracking it correctly to see changes in behavior.

8.3.1 Collecting Data

Many businesses have been collecting data for years and years. They have employees who oversee this data and know how to analyze it. Still, it is not guaranteed that the current data collection is adequate regarding the development of the recommender system. Changes may have to be made with respect to how the data is being withdrawn from domain activity. There is also a possibility that the targeted domain does not have any established data collecting methods. It is therefore a crucial step in order to make sure that the targeted domain has the adequate data collection measures established in order to proceed with the development of the concept.

What all recommender systems have in common is the need for a strong understanding of the affected users, or at least a strong understanding of the data that is needed from the users. Therefore, it is very important to investigate and define the data collection as either implicitly (recording behavior and patterns) or explicitly (collecting data by inquiry) (Katakam, 2019). The reason why this is important, is to make sure that data collection aligns with the chosen recommender system.

8.3.2 Structuring Data

It is not enough to just collect data if there is no structure to it. The structuring of data is what creates useful insights from interaction on the domain, which is what the recommender system's logical reasoning is based on. Implications of new knowledge can therefore be spotted with the right data structure. Structuring data is closely related to collecting data in that it is a part of the overall approach to establish a holistic usage of multiple datasets. It is important that the data in question can actually be analyzed and conclusions can be drawn based on its patterns. This is especially crucial when it comes to recommender systems. For example, content recommendation may need in-depth metadata tagging of both users and content to function properly. This is a best practice in some of the biggest media companies, as mentioned earlier. Following this example, logical implications from proper metadata tagging may be the result of a useful data structure. This is also a reason why it is important to work with data semantically, as when a semantic layer is used. Regarding the concept at hand, the filtering approach for the recommender system should already be decided upon when the company's handling of data is examined. As such, it is the demands of the filtering approach that defines if the company's structuring of data is sufficient, or not, for the recommender system.

8.3.3 Tracking Data

It is important to be alert of shifting user interaction patterns. With the proper tracking and overseeing of data, it is easy to spot how users might start to change their behavior. Gaining this knowledge can give the business the time needed to re-adjust to the behavioral changes, optimizing their strategies accordingly. It is very important to properly track data when the concept goes live, since it may provide knowledge on whether users find the recommendations useful or not. As mentioned earlier, recommendations can be invasive in the user experience, therefore, it is valuable to track just how much interaction they provide in the bigger picture of the business domain to make sure that users find them useful. Proper tracking of data also provides a foundation for experimentation and testing of new features in a controlled manner, since irregular implications in the data can incentivize optimization accordingly.

All of these three aspects of data handling are crucial to examine in order to optimize the conditions for establishing a semantic layer. With all three accounted for, it becomes possible to start creating a safe basis for optimal content recommendations in practice. This raises the question of how to map the relationships between data, to argue for semantic interconnections of information that bring forth recommendation possibilities to the end user.

8.4 Step 3 - Formal Ontology

Now that all required knowledge about the domain has been gathered, and necessary decisions regarding the recommender system have been made, the development of the ontology (which establishes the semantic layer for the chosen information environment), can commence. In reality, implementing a recommender system would already be possible at this point. Information architecture has already played a vital role in the development so far (analyzing and mapping the affected domain with the use of the Information Ecology Model). However, it is the semantic layer that creates the insightful value aspect of this concept in comparison to automated machine learning principles.

"an ontology is a formal explicit description of concepts in a domain of discourse (classes*), properties of each concept describing various features and attributes of the concept (slots), and restrictions on slots (facets of slots)."* - Noy & McGuiness, 2001.

By formally describing the different elements of the information environment as concepts (classes and individuals), properties (slots), and facets of properties in a given domain, the domain can be semantically structured. This mapping allows for the making of explicit assumptions regarding the domain. It also provides a shared understanding of the domain's logical structure. Therefore, in the context of building the optimal recommender system for a given domain, the ontology is a valuable tool. By enabling the use of the complete knowledge base of the domain, more precise and personalized recommendations can be defined. The practical value of the ontology is seen in its capability to answer competency questions regarding the domain. By examining the contextual knowledge of classes, individuals and their properties logical patterns should occur, providing the answers to established competency questions. As an example, in the case of a recommender system, the ontology should most likely be able to provide answers for who to recommend specific pieces of content to, and what to recommend. This depends on what the established competency questions entail.

8.4.1 Ontology Development

The creation of ontologies is often an iterative process. This is also the case with the ontologies that are to be based in this concept. The early phases of the development of the ontology will follow the approach of Noy & McGuiness and base the decision-making on formerly established knowledge, gained through the steps of the concept. It is when the ontology is to be implemented, that the iterative aspects of the process will become apparent.

8.4.1.1 Define Domain and Scope

As Noy & McGuiness state, the first step, when developing the ontology, should be to define the domain and scope. In order to define this, specific questions form the framework for what this step must answer. What (domain) will the ontology cover? What is the purpose of the ontology? What are the competency questions that the ontology must be able to answer? Who are the individuals or departments who will operate and maintain the ontology? The consultants should already be able to answer most of these questions by having established the domain and the purpose of the recommender system in *Step 1* of the concept. The domain and purpose of the recommender system will in most cases correspond well with the domain and purpose of the ontology. The competency questions, which the ontology will provide answers to, should be created based on the established domain, scope, and purpose of the ontology. Like with the domain and purpose of the ontology, the conversation of who should operate and maintain the ontology should already have taken place in *Step 1* of the concept, and as such, the answer to the question should be fairly obvious to the ones responsible.

8.4.1.2 Examine Existing Ontologies

According to Noy & McGuiness, the second step of the process should be to examine and use any relevant existing ontologies. As different companies are expected to have very different standards for their knowledge, experience, circumstances, and capabilities regarding ontologies, this step will differ greatly depending on the company and domain. If relevant existing ontologies are eligible for use in the development of the new ontology, the use of existing classes, individuals, and properties should be considered for use.

8.4.1.3 Enumerate Terms

The third step of the development process consists of the enumeration of terms for use in the ontology. The individuals and classes of the ontology are all to be based on appropriate terms that exist in the given domain. Therefore, an extensive enumeration of terms is needed. There are many different methods for enumerating these terms. Based on the size of the domain, different measures may be more effective than others. A classic brainstorm and listing may suffice in some cases, while other cases may require more strategic tools like with the example of a content audit (Hedden, 2010). Depending on the OWL-editor that is used, it may also be possible to do the enumeration of terms directly in the editor program, rather than creating a separate listing.

8.4.1.4 Define Classes and Class Hierarchy

The fourth step of the Noy and McGuinness approach states that classes and a class hierarchy must be defined. As formerly mentioned, the definitions of the different classes are based on the enumerated terms. The challenge of defining classes from the enumerated terms stems from the terms being able to be defined as either classes or individuals based on the context of the domain. As such, the domain and purpose of the ontology must be considered when defining classes.

There are many different ways of establishing a class hierarchy. First, the architecture of the hierarchy needs to be defined. Common architecture-types are the classical top-down, bottom-up, or a combination hierarchy. As this thesis is expected to use the OWL-editor Protégé, the architecture of the class hierarchy is likely a top-down hierarchy. Defining the hierarchy's super- and subclasses comes next. This may be the challenging part of the process, depending on the complexity of the domain. When defining the status of the class in the hierarchy, the quote "If a class A is a superclass of class B, then every instance of B is also an instance of A" (Noy & McGuinness, 2001) should be considered.

8.4.1.5 Define Properties

With an established class hierarchy, the fifth step entails properties of classes, which are to be defined. Firstly, the properties must be defined as either being:

- Object Properties (linking individuals to individuals)
- Datatype Properties (linking individuals to data about the given individual)
- Annotation Properties (add information about classes, individuals, or object/data type properties)

When the type of properties has been defined, inverse properties can also be defined, if relevant.

8.4.1.6 Define Facets of Properties

At the sixth step, the different facets of the established properties should be defined. The required facets may differ greatly depending on the ontology at hand. It is at this step that, most commonly, the cardinality of the property, the value type of the property, and the domain and range of the property will be defined. It is also at this step that it makes sense to define the needed restrictions of the properties. As such, the decision to introduce either *quantifier* restrictions, *existential* restrictions, and/or *hasValue* restrictions must be made. The domain and scope of the ontology will likely play a large part in the need for facets of properties.

8.4.1.7 Define Individuals

The seventh and final step of developing the ontology, according to Noy & McGuinness, consists of defining the individuals out of the classes in the hierarchy. With the earlier steps completed, the individuals should contain a lot of information about them and their relationship to other individuals as well as other classes.

8.4.2 Recommendation Filtering Logic

Depending on the chosen filtering approach, the overall semantic structure of the ontology should be able to change accordingly. Examples of how this is expected to be accomplished for the most prominent filtering approaches are explained here.

As an example, with the *popularity filtering* approach, different classes or individuals, consisting of content, will earn a score based on the interest they have been shown by users (likes, conversations, views etc.). As such, the ontology needs to describe the interest that is shown to the classes or individuals. This can be accomplished with *object properties*, as they will allow the ontology to show the logical relationships between content and users. Through these relationships, content will be shown to contain interactions (highest number of likes, largest amount of interactions) and can therefore be defined as the most popular content to be recommended.

When it comes to the *collaborative filtering* approach, as stated, the focus is on the users. Therefore, it is the classes and individuals regarding or consisting of users where *object properties* become valuable to the ontology. Through the relationships between users and content, patterns in users' interactions and behavior will become apparent, allowing the algorithm to base its recommendations of content on these patterns.

For the *content-based filtering* model the focus is on the content. As such, the classes or individuals, pertaining to content, need *object properties* to establish similarity through relationships between content. The patterns of similarity, or dissimilarity, are what the algorithm will base its recommendations on.

As explained, the *hybrid filtering* approach combines multiple models. As such, the individuals and classes of the ontology demand a greater use of *object properties*. Depending on the chosen filtering approaches for the *hybrid filtering* model, relationships of both classes *or* individuals regarding both users and content must be defined. The level of descriptive detailed mapping of the elements of the ontology may differ depending on the purpose and domain of the ontology.

8.4.3 Implementation in Protégé

Having accounted for the design of the ontology, it is now time for implementation in Protégé. This is very important since Protégé is rooted in the OWL-language. With this, it is easy to utilize the implementation to answer competency questions and ultimately argue for noteworthy findings that can be used for deciding on specific recommendations. Protégé is an editor based on logic. The concept relies on logic for the argumentation to say that a certain user needs, or has an interest in, certain content, and thus sparks the possibility of a recommendation based on data patterns alongside a logical reasoner attached to the program. The idea is, fundamentally, that by taking the established classes, individuals, and properties from the development stage, and plotting them into Protégé with the correct definitions (based on the OWL-language), it becomes possible to gain insights and answers to initial competency questions and curious patterns in this structure of semantic information.

With the right guidance, the people responsible for the domain will most likely be able to implement the newly created ontology into the business structure and utilize it in order to create the most optimal recommendations on their domain. They might even expand their ontology to other related subdomains, and thus utilize semantically structured data for multiple purposes across a broader business domain. A result of this could be new algorithms situated on the same semantic layer, and result in things such as cross-recommendations from different information environments. For example, this could lead to bespoke package deals of content, which utilize knowledge from multiple domains of interaction.

8.4.4 Evaluation of Competency Questions

With the full development of the ontology having been completed, its ability to provide answers for the established competency questions can be evaluated. By examining the ontology's explanation for relations between the different elements of the domain, interpretation of those explanations will bring forth answers to the competency questions. By being able to provide answers to the competency questions, the practical applicability of the ontology is showcased, and the value of semantically structured data for the specific case is expected to emerge. This is to be discussed in detail with the client. This ties back to the *collaborative mindset* of the *Concept Model*.

8.4.4.1 Interaction Graph

Besides the implementation of the ontology in Protégé, an interaction graph will also be created (Figure 16). For example, this was the case for Pinterest, after they had developed and implemented an ontology, in order to showcase its basic use. For this concept, the interaction graph is more inclined to be a conceptual view of the recommender system and how it ties into the ontology. The purpose of the interaction graph is to showcase the logic of the recommender system, and how it is able to provide answers for prominent competency questions. In practice, the interaction graph will give the people responsible for the domain a clear understanding of the purpose and usage of the recommender system in conjunction with the logic of the ontology.

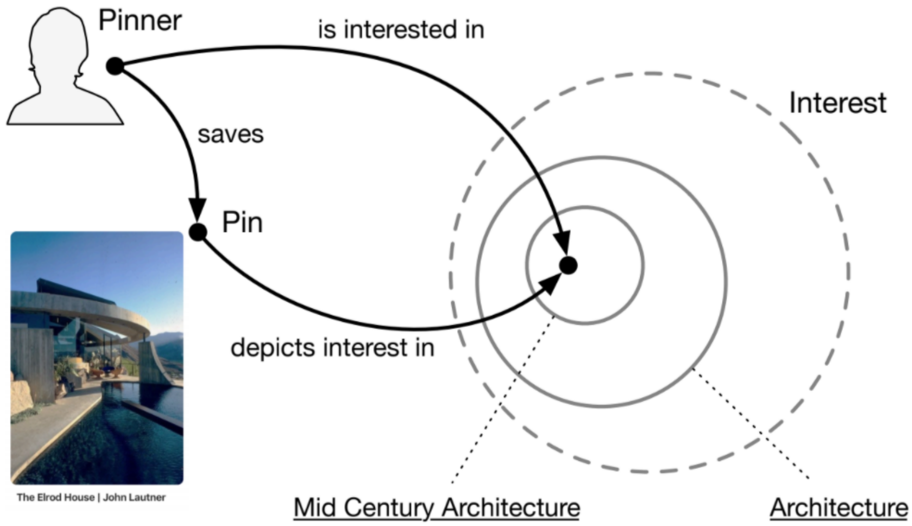

Figure 15: Conceptual view (interaction graph) of Pinterest Taste Graph (Gonçalves et al., 2019).

8.5 Step 4 - Persuasion

Having finished up the implementation of the ontology in Protégé, it is finally time to utilize some of the aspects of persuasive design in the concept. This part is important, since it is what accounts for new knowledge compared to the existing recommendation systems that have been implemented into ontologies as discussed in the *State of the Art* section of this thesis. Utilizing persuasion is an important factor in creating the communicative aspect on top of the arguments that the ontology is able to provide. If a business wants a recommender system that is ultimately able to tell what kind of information to recommend and *how*. The '*how*' consists of persuasive principles that will assist in the arguments for how to communicate and visualize a given recommendation.

8.5.1 Implementation of Persuasion

To answer the '*how*', in terms of this concept and the recommendation capabilities herein, a set of persuasion principles have been chosen to start establishing possible outcomes. An example of a persuasion principle for this concept is the Behavior Model (Fogg, 2009). In concept, the model can be utilized for gaining insights into a particular user's personal interaction capabilities in order to interact with recommended content. This is specifically in terms of their *motivation* or *ability* in the specific situation. The thing that makes the model suitable for a concept like this, is that it will likely fit many business domains because of its initial simplicity. The model will not go in-depth with the user's psychology, but will instead touch on some of the persuasive aspects concerning the user's superficial reason to act on recommendations. The goal of acting on these aspects is to give the user more reason to interact with a recommended piece of content, by showing them the right information that supports this reasoning. This is done through a *prompt*, which assists in establishing the right approach to communication (rooted in either *motivation* or *ability*) for a recommendable piece of content. The simple nature of the model also makes it easily attachable on top of the existing semantic groundworks that have been laid by the ontology.

The "toolbox" of persuasive principles should expand as the concept becomes more refined through practical usage, giving the established concept a possibility to become more advanced as the theoretical approach is being tested and utilized. If the use of the Behavior Model proves successful, it increases the likelihood that more similar persuasion principles will join its ranks and be utilized in the concept. This has already been touched upon in the *Limitations* section of the *Methodology* chapter. It is important to stress that persuasion is difficult to define as a specific reusable thing that will act the same for all concepts. As long as persuasion keeps proving successful as a facilitator of the concept's icing-on-top aspect of establishing communication for evaluated recommendable content, it is worthwhile to investigate in-depth for each individual concept in order to find the perfect principles.

8.5.1.1 Risk Awareness and Persuasion

As mentioned in *Step 1*, risk-aware recommender systems often ensure that the time, place, amount and presentation of the recommendations are taken into account – what could otherwise be called the persuasiveness of the recommendations. An example of the different needs of risk awareness and persuasion in recommender systems could be tied to the choice of platform. A recommender system for a web domain might not need the persuasion principle of Kairos (Glud, Jespersen, 2008). If the recommendation is already stated to be placed statically on a site, then the specific time and place, i.e., Kairos, becomes redundant. However, a recommendation system for a mobile application domain might benefit from the use of Kairos. If the recommendation has already been determined to exist as push-notifications, then Kairos could assist in planning the opportune time and space to provide the user with the recommendation on the go. As stated, an approximation of how risk-aware the recommender system should be is defined early in the concept process, while the practical design is decided upon on this step.

8.5.2 Evaluation of a Practical Recommendation

After having implemented persuasive principles into the recommender system, it is now possible to evaluate an actual content recommendation, and thus the theoretical concept in its entirety. This is to be done in a way that can give meaningful information to the business about their recommendation-domain in question. This step will give a visualized idea of how an actual recommendation should look and behave in a practical setting, be it on a website, inside an application or through push-notifications. It will furthermore account for rhetoric in communication, as this often ties into the persuasion aspect.

8.6 Insights/Maintenance

This post-concept development step will require a conversation with the business about insights and maintenance possibilities. It is not automatically assumed that the business will have a complete grasp on how to utilize the concept after delivery, even though they have been included in the work process of many of the steps, as shown with the *collaborative mindset*. This is why it is an important matter to discuss, so that the process of implementing the concept into the business structure will be as smooth as possible.

Consultants can also reflect upon insights from the process as a means of strengthening the approach for the next concept development. Since each concept is different, there will undoubtedly be a lot of insights that are made with each new collaboration. In this regard, it is important to evaluate both good and bad things that happened during the process. After having established a distinct routine with the concept, it will be possible to experiment with some of the steps of the concept, to see how different approaches to certain problems might be more adequate.

If the concept is to be developed in a consultant practice, and the collaboration stops, it may be beneficial to contact the business again after some time, to investigate how they are maintaining the solution. This gives insights into how the company has either successfully or unsuccessfully maintained the solution, which might also give certain insights into future collaborations and give points of discussion to include in new company meetings and the like. In professional settings, available time is often quite limited, and therefore it is not always easy to find time for reflection. However, in the long run, it is worth building in time for a reflective stage of the process, especially if the practical solution has not been utilized outside of a theoretical stage for a substantial amount of time. Therefore, the more theoretically rooted the approach is, the more important it is to reflect properly.

As the walkthrough of the theoretical concept comes to an end, it is important to include its academic relevance in terms of what the concept can theoretically accomplish for the initial hypothesis of the thesis. The concept creates a practical product, which is expected to be usable for a large number of businesses. This product has a role in showcasing the practical value of information architecture, specifically of semantically structured information. This is specifically related to the value of logical insights, gained by deploying a semantic layer for recommender systems. These perspectives all tie into this concept, and how it is expected to function as a consultant endeavor for data-driven businesses. It is now important to move from a theoretical approach to a practical approach by utilizing the concept on two real data-driven companies.

9 Introduction to Practical Cases

Having established the theoretical concept, it can now be utilized in practice, to showcase its capabilities. The theoretical concept is to be utilized as a framework for practical cases.

The concept is meant to be usable for many companies of different organizational maturity, resources, and contexts. To ensure that the concept is relevant for the companies in question, and that the defined principles of information architecture become an integral part of the practical cases, specific criteria for the chosen companies must be met. The chosen companies must be data-driven or plan to become data-driven, meaning they should have, or have the potential to acquire, a large amount of data for semantic structuring. For this, they need to have some type of customer/user and produce some type of content, as the idea of the recommender system is to recommend content to users. The company also needs to be present in the digital world via platforms, websites, apps, or other digital spaces. For this thesis, two practical cases have been chosen for the theoretical concept to be utilized in. These two companies are OpdagDanmark and NNIT.

9.1.1 OpdagDanmark

OpdagDanmark is a startup-business from Aalborg, Denmark. The core idea behind OpdagDanmark is to transmit information about experiences around the country, be it hiking trails, camping spots or just a good place to get a pizza. They then make articles and guides to get people motivated to try out these experiences, and plug the experiences into an overall experience map on their digital platforms. At the time of writing, the utilized digital platforms are their website and an application for smartphones, which is still in development. This experience map helps to give an overview based on the location of the different experiences available (with guides and articles), and offers a nice explorative user experience that is set up to inspire users to try things they may not have otherwise considered.

The way this idea differs from standard tourist organizations (who are restricted by sponsorships) is by giving the users the power to judge what is worth plugging into the map, by organizing polls that are run by an internal SMS-voting system. The winners are given the honor of "*Best Pizza in Denmark 2020*" for example, which is then displayed on their designated article (located on the experience map). Ultimately, all places in a given experience category should become mapped at some point in time, giving an overview of everything inside that category that is offered in Denmark, but only the ones deemed worthy by users will receive awards or similar accolades. This helps brand the places that the users deem the absolute best, which is the basis of the concept that is offered; helping users to find the "hidden gems" within particular categories.

OpdagDanmark is only run by two people, but the company is starting to have a fairly large number of active users. This is especially the case for their application. Currently, the total number of app downloads has exceeded 15.000, and it is still rising. The real problem that OpdagDanmark is starting to face is that their usage of data is not optimized properly based on the possibilities they have to hand. They simply do not fully utilize the data that they are capable of collecting from their users. Because of this, they function as a good practical case for the thesis and the established concept.

OpdagDanmark's smartphone application is chosen in order to investigate the practicality of information architecture in a small startup-business (with a lot of potential for utilizing data about users and content). The established concept could be what OpdagDanmark needs in order to strengthen the user experience of their application, by giving users personalized experience recommendations that are tailored to their individual tastes. This has a lot of potential regarding the use of an ontology. The logical implications and general knowledge about user-content interaction that can be gathered, combined with the persuasion aspect of communication recommendations, will allow OpdagDanmark to optimally tailor a satisfying experience for their users. Optimizing the user experience will hopefully inspire the users to experience more, relevant leisure activities around Denmark.

9.1.2 NNIT

NNIT is a global IT-service provider. The company is based in Denmark and is one of the country's largest IT-service providers with more than 3.000 employees. The company's primary goal is enabling their customers' digital transformation. The company accomplishes this by consulting, developing, implementing, and operating modern IT solutions and services. NNIT provides these services in many different industries, including Life Sciences, as one of their largest industries. The company also operates in Logistics, Finance, Retail, Industrial products, the Private and Public Sectors, Energy and Utility, Consumer products, and Healthcare.

With a company as large as NNIT, the opportunities for introducing value-creating recommendation systems are innumerable. In order to make use of the established concept in a way that will produce practical value through the designed recommender system, the domain for the recommender system has to be very specific and comprehensible. To assist in defining this domain, collaboration between consultants and NNIT's *Communications and Marketing* department (*CoMa*) will be established. Even with a focus on the *Communications and Marketing* department, the domain is still too vast. The CoMa-department has many subdivisions and platforms that it is responsible for. The department is divided into *Marketing*, *Press and Media*, and *Internal Communication*. The department is also responsible for NNIT's commercial website: nnit.com. This website is primarily part of the *Marketing* and *Press and Media* divisions, as the goal of the website is to provide commercial knowledge about all of NNIT's services, industries, and external endeavors. The website has visitors looking for information and services, while the CoMa-department produces different types of content that are presented on the website hoping to assist visitors in their purpose for visiting. This means that, while the company in question is NNIT, it is the CoMa-department and specifically the information environment of nnit.com, that is the second chosen case for this thesis. With this thesis' resources in mind, the whole of nnit.com is still too broad a domain for the established concept. Scaling down the domain further, and finding a very specific domain on nnit.com for the recommender system, will be accomplished in collaboration with the people responsible for the CoMa-department.

An overarching recommender system for the whole of NNIT.com would, of course, be expected to bring more beneficial results to the CoMa-department, rather than a small specific subdomain of the website. However, it is important that the concept is scalable depending on the needs of the client. As such, the recommender system that is to be built for the small subdomain, is to be viewed as a proof of concept that, with more resources, could be scaled-up to encompass all of nnit.com. The insights that can be gained through the concept are also expected to be valuable for the client. Through the deployment of semantic layer, the findings of the case are expected to assist the client in a more valuable utilization of their data.

9.1.3 Initial Comparison

So why have these two companies been specifically chosen? How will these cases, when their findings are compared, be able to provide answers to the thesis' problem area and hypothesis? It is important to emphasize that the findings and comparison of the two cases will not be able to provide definite answers to the questions and problems presented by the thesis. For these findings to be definite, the established concept should be applied in practice, and tracked over time, for a wide range of different companies. However, the two practical cases of this thesis are expected to create a foundational understanding of the practical value that can be gained by utilizing data semantically (in this case for the purpose of developing a competent recommender system). With the chosen companies being widely different in their purpose, organizational maturity, and available resources, the cases are also expected to showcase how the established concept is able to provide differing levels of value depending on the domain. This will give an in-depth contrast, and a multi-dimensional perspective, that will be able to pinpoint strengths and weaknesses of the concept in a practical setting. This thesis does not explicitly target any particular type of company, as long as they fulfill the defined criteria. Therefore, with the perspective of solving a problem for many different companies, if the two chosen companies were similar, the findings of their cases would provide very shallow knowledge. By choosing companies that are distinctly different, it is expected that a more profound understanding of the practical implications of utilizing data semantically can be obtained.

10 Practical Case for OpdagDanmark

10.1 Initial Meeting

To start off the collaboration with OpdagDanmark, a meeting was held between the consultants and the two employees of the company. It was fairly easy for the company to understand the core idea of the concept, and what kind of value it would be able to provide. At this initial meeting, the overall purpose of the concept was established alongside the chosen information environment (where the recommendations should be implemented to create the most value for the users).

The purpose of recommendation for OpdagDanmark is very straightforward. Recommendation is an interesting addition to the existing application as it perfectly fits with the explorative philosophy of the company. OpdagDanmark is specifically interested in a deeper recommendation service, which is not only developed based on previous user-content interaction, but also on the time and place of the specific user (essentially like the Kairos-moment). That is where they envision the best use of the concept idea on their domain. This is also a good way to test the logical and persuasive aspect of user recommendation on a smartphone application with all its possible recommendation options.

For example, if a user was hiking on a route, listed in the app (functioning as a piece of content), the app should be able to recommend a nearby place to eat afterwards (another piece of content), which also matches the preferences of that specific user; essentially creating a really strong content recommendation. OpdagDanmark is interested in having this recommendation option communicated to their users as a push notification, since this could raise the inspirational guiding potential of the app.

To establish relevant recommendations like this requires a lot of data about the user, which is something OpdagDanmark has carefully collected since the release of their application. From the very beginning, they heavily focused on the collection of user data while developing the framework of the application. They knew that, eventually, utilizing data would assist in developing a personalized and in-depth user experience. This attention to collecting data gives the concept a head-start. Even though the app is in its early stages and not developed fully, there are sufficient amounts of data to create a functioning template of a recommendation service with the existing collection of data and future potential therein.

The specific data that OpdagDanmark has collected includes data about phone operative systems, voting history in the polls, and coordinates (location data). This is confidential information for the individual user and will remain confidential in the process of writing this thesis. It is therefore important to stress that the data used in the later sections of this development process is anonymized and not based directly on the confidential existing user data. This is important because of the confidentiality breach that would occur, if this thesis is published with this information present. However, since it is only the data collection pattern that is important and not the actual data itself, this course of action will assist in making a proof of concept with similar, but fabricated, data.

10.2 Step 1 - Business Analysis

10.2.1 Ecology Model

As the concept has mentioned the Information Ecology Model as the initial starting point for analysis, this will be utilized for a holistic approach to how OpdagDanmark should implement a recommender system in practice.

10.2.1.1 Context

First of all, it is important to look at the context of the overall recommendation domain. In OpdagDanmark's case, they find it useful to be able to provide a personalized recommendation about a single experience that is physically located close to the receiver, which is then communicated through a push notification.

A push notification is essentially a notification on the home screen of your smartphone (regardless of operating system), located in the same place that a smartphone user receives messages from friends or family, or alerts if their battery is running low. These kinds of notifications help to boost the explorative purpose of the application, and make its guiding capabilities stronger, since it becomes helpful outside of the direct app experience. However, it is risky to notify users in this way, and is best suited for specific moments where a lot of data backs up the fact that the user will find a specific recommendation useful. If the user does not find the recommendation useful, they might become annoyed with the application, resulting in a possible uninstall.

10.2.1.2 Content

What kind of content is supposed to be recommended in the push notifications? OpdagDanmark's typical content consists of articles and guides about experiences that can be explored by users when using the *experience map* (Figure 16). The content that should be included in the push notification could be a specific experience that would benefit the user in a particular situation. There are many different types of experiences like hotels, restaurants, hiking trails and much more. The different types of experiences also have different levels of potential for being interacted with, when being recommended through the notifications.

For optimal recommendations, the user should be near the experience they are getting notified about and have a certain amount of motivation for a given experience at that moment (dinner time for a restaurant experience, for instance) such that it is not a problem to persuade them to try the experience. Experiences like restaurants can have a stronger potential for success, as the recommendation can be spontaneously accepted. This is in contrast to experiences like hiking trails, which usually require a certain level of planning for the user. Focusing on push notifications, it is possible to rank the experiences in terms of inclusion into this recommendation domain.

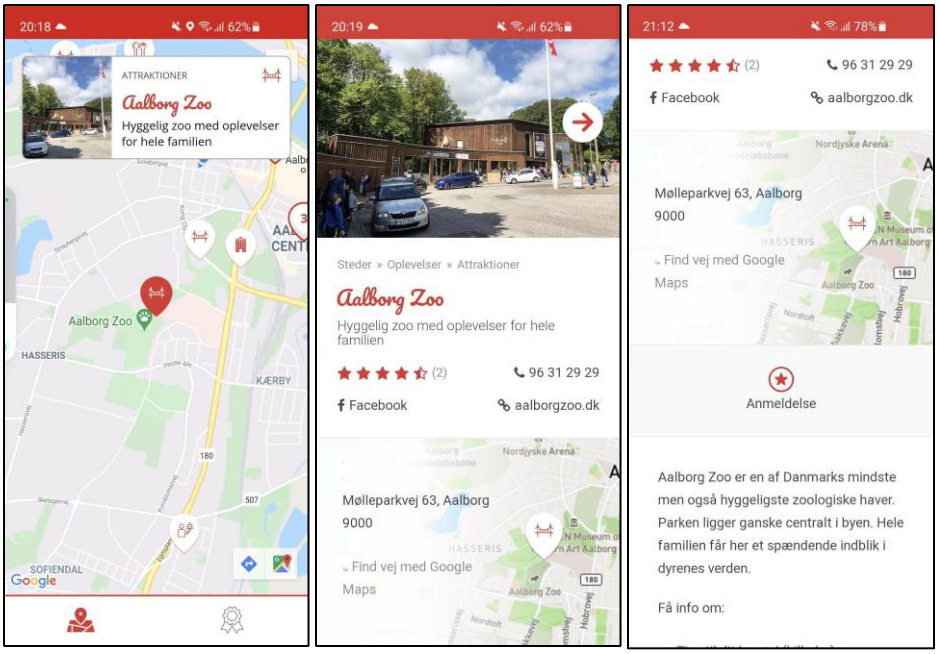

Figure 16: An experience on the experience map with the associated article – OpdagDanmark.

10.2.1.3 Users

Based on data from social media that OpdagDanmark has provided about their users (Figure 17), it is interesting to note that a large proportion of them are women in the age group 45-54. Based on further analytics data, OpdagDanmark has confirmed that this is also the most common type of user that currently downloads and utilizes the application. This model therefore shows the most likely demographic distribution from the array of anonymized data strings that OpdagDanmark has acquired, and which can be utilized in the concept. This is not a certainty, however, since the domain does not require login information (yet), and users are therefore only logged as numbers.

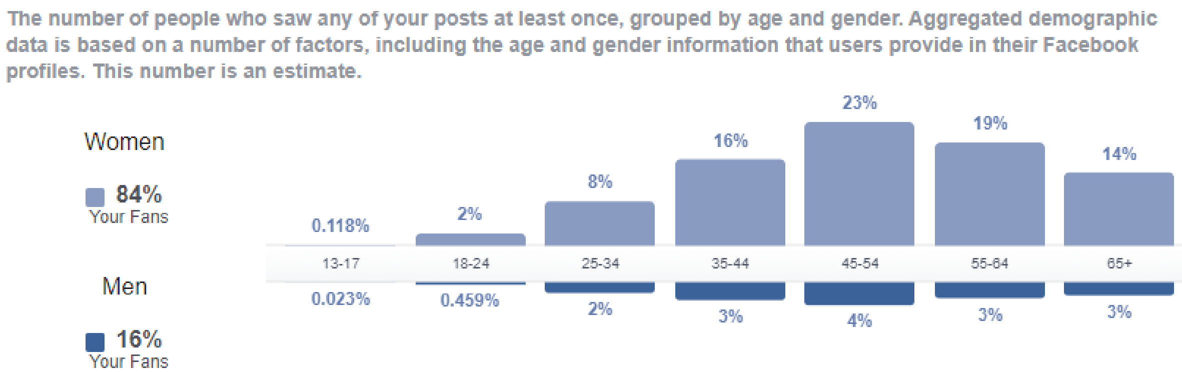

Figure 17: Data from Facebook-insights (34,000 followers) – OpdagDanmark.

As most users are defined as middle-aged women, it implies that this type of user is typically the decision-maker of a larger family (Holst, 2016). Therefore, they are the ones in charge of holiday destinations among other things. They are likely using the app for inspiration for where they want to go, and then scheduling it into an overall vacation plan. The personalized push notification does not directly match with this type of user. Recommendations through push notifications generally support a more spontaneous approach to experiencing Denmark. The defined user, in contrast, possibly takes a more considered approach, and plans experiences beforehand, using the spontaneous push notifications as a tool for inspiration rather than something to act upon immediately. Initially, this poses a problem for the value that can be gained with recommendations through push notifications.

However, there is another perspective to the implementation of push notifications for OpdagDanmark. This kind of recommendation might be more valuable for other types of users and incentivize a more diverse demographic on the application in general. This could be an interesting thing to investigate alongside the general value perspective of recommendations through push notifications, as OpdagDanmark are very open to addressing user groups that they do not yet have a significant quantity of. This is the case with young adults, both male and female, particularly in the age group 18-24, who are almost non-existent in the visualized data. If this demographic became more interested in the app, it could result in a large value aspect for OpdagDanmark. This could help even out the demographic by attracting users from more age groups to download the app.

Implementing the right type of push notification could therefore become a step in the right direction for catering to this “new” type of user, and this could hopefully assist in smoothing the demographic distribution on the application. An important consideration is that users from a younger demographic are more likely to be in a situation where they can act upon OpdagDanmark’s recommendations, in contrast to middle-aged women who, as stated before, tend to plan their trips in detail because of the family aspect. The younger demographic might therefore be more receptive to these recommendations.

10.2.2 Evaluation

Based on the information that has been analyzed from the three subdomains, it is worth investigating the idea of push notifications as the overall domain for the recommendations. This matches the investigative approach of catering to a new demographic, which is also something OpdagDanmark has shown interest in exploring themselves. OpdagDanmark currently has some experiences on their domain that are targeted towards a younger demographic *on the go*, meaning they are often traveling and physically moving between places. This mostly involves food, since it is important that the featured experiences are spontaneous experiences, and not experiences that require booking in advance etc. As such, it may be reasonable to focus specifically on fast food since this experience does not require much planning from the user and is therefore more likely to be persuasive as a recommendation (in terms of motivation and ability aspects regarding time and money, for example). OpdagDanmark has plenty of fast food experiences in their catalog of experiences, so this is a good starting point for the purpose of making persuasive recommendations.

Location-wise, it would only make sense to notify about experiences close to the location of the receiver. As some experiences are located in busy metropolitan areas, their chance of being able to become a recommendable piece of content rises due to increased footfall. The starting criteria for the push notification are therefore spontaneous experience types, that will typically be located in larger cities, so that they are recommendable more often. OpdagDanmark has most of their experiences listed in the larger cities of Denmark, so this is also something that is easily manageable with the currently available information.

Regarding the user’s situation when receiving the notifications; If the younger demographic approaches the app with an *on the go* type of perspective, to locate experiences in a new environment that they are visiting, they may be more easily persuaded and gain more value from recommendations through push notifications. In familiar environments, such as the user's current city of residency, it is possible that the user might already be familiar with an experience that is recommended to them. It is therefore optimal that the user does not receive recommendations in these kinds of environments, as the likelihood of them becoming annoyed will probably rise in conjunction with the aspect of familiarity; at least if the recommendations are perceived as an exploratory feature, which is the intent.

Familiarity is not something the application can consider, since it does not have data about where and when the user has visited certain places. If the app had detailed information regarding the user's interaction history with (as an example) fast food places over an extended period of time, then it would become an option to recommend experiences in the user’s hometown. However, in terms of an approach to recommendation for this particular case, the focus must remain on new experiences, and as such, experiences that are not located close to the user’s residence. This is also a result of what is realistic and valuable to pursue with the current organizational maturity of OpdagDanmark. In any case, expanding the recommender system’s capabilities by being able to provide deeper, more reliable, and personalized recommendations (with both familiar and unfamiliar environments) is an opportunity to increase the future capabilities of the system.

In summary, the type of recommendations is *push notifications* (something that OpdagDanmark is very interested in implementing). These notifications should include *spontaneous experiences* (most likely fast food places, since these can be considered easier to persuade the user to try) for a *younger demographic* (since people aged 18-24 are almost non-existent on the application currently). This demographic should also be *on the go* (physically moving around in an unfamiliar environment).

10.2.3 Recommender System

In order for valuable push notifications to be delivered to relevant users, it is necessary to conclude how OpdagDanmark's existing (and obtainable) data will function in collaboration with filtering, and how the circumstances of the system affect its design.

10.2.3.1 Deciding on a Filtering Approach

The filtering approach is largely based on the amount and type of data that is available. As there is no current deep understanding of the interconnected behavior between the user groups of the information environment, *collaborative filtering* would be met with the *cold start* problem (Sheridan, Onsjö, Becerra, Jimenez & Dueñas, 2019), and would therefore not be optimal. Instead, it would be beneficial to look into *content-based filtering*, where the content that a given user has interacted with, is what determines the recommendation that they will receive. This is, of course, only if the user has indicated that they have an interest in the content.

In order to establish a starting point for this type of filtering, it makes sense to look at certain data regarding what kind of content users have been voting on in *Denmark's best* polls. These polls are available on both the web domain, but also the application, where the *user_id* (the specific number-based ID associated with a user that has downloaded the app) is linked directly to a vote in the poll-system (Figure 18). There is information about date and time of voting, the user's individual ID-string, the title of the overall poll, and lastly the participant (the experience being voted on) alongside the participant's address. These data strings do not breach confidentiality, as they are based on test accounts.

date	user_id	voting_title	participant	participant_street
17-02-2020 14:05	10000	Danmarks Bedste Badehoteller 2020	Strandgaarden Badehotel	Strandvejen 8
19-02-2020 08:51	10043	Danmarks Bedste Badehoteller 2020	Strandgaarden Badehotel	Strandvejen 8
15-02-2021 12:04	10043	Danmarks Bedste Badehoteller 2021	Strandgaarden Badehotel	Strandvejen 8
15-02-2021 23:54	10000	Danmarks Bedste Badehoteller 2021	Strandgaarden Badehotel	Strandvejen 8
17-02-2020 14:05	10000	Danmarks Bedste Badehoteller 2020	Villa Vest Badehotel	Rubjergvej 2
17-02-2020 14:11	10043	Danmarks Bedste Badehoteller 2020	Villa Vest Badehotel	Rubjergvej 2

Figure 18: Extract of dataset for app-users voting for experiences in polls – OpdagDanmark.

This data indicates certain users' voting patterns, which directly translates into what kind of content they are interested in. In other words, by looking at the users' voting patterns for the different yearly polls held by OpdagDanmark, conclusions can be drawn as to what kind of experience each individual user likes. More importantly, it can be utilized to shape a deeper understanding of the user's preferences in terms of the specific experience type. For example, if a user votes on a pizza place (defined by being any sort of establishment that serves pizza) for "*Denmark's best Pizza 2021*" the specific preferences for the experience category "*Pizza*" are now linked to the user and used for profiling. These are things such as using, or not using, organic ingredients, general price range, or the overall selection size etc. If the user is then traveling to a new city, they can receive a recommendation for a pizza place similar to the ones that were voted for, as this fits their preferences. This can then be considered *content-based filtering*, as the recommendations are based on the positive interactions with prior experience (by voting in the poll).

10.2.3.2 Circumstances of the Recommender System

Since the recommendations are supposed to take time and place into consideration, it is important to have data that backs this up as well. Because of the nature of push notifications, the recommender system must be risk-aware and be able to consider the circumstances of a specific recommendation before actually sending it. This gives more layers to the algorithm that must also be backed up by relevant analysis of data to make sure the circumstances are optimal. For this situation, OpdagDanmark has datasets that sufficiently back up the risk-aware nature of the system (Figure 19). These datasets are based on when a particular user opens the OpdagDanmark application on a smartphone. There is information about the user's individual ID-string, the ID for their device, the latitude and longitude of their current location, and finally, date and time of access. This data is once again based on the same test accounts as the previous dataset, and does not breach confidentiality.

user_id	device_id	latitude	longitude	created_at
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.033264842	9.907808271	12-03-2020 20:22
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.033264842	9.907808271	12-03-2020 20:41
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.033264842	9.907808271	12-03-2020 20:59
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.033264842	9.907808271	12-03-2020 21:07
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.036568140	9.909373150	12-03-2020 22:35
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.036568140	9.909373150	13-03-2020 07:05
10000	08b76530-120e-4f94-aae4-df9e191753cd	57.036568140	9.909373150	13-03-2020 07:11

Figure 19: Extract of dataset for tracked locations based on app usage – OpdagDanmark.

This data can be used to reinforce the risk-aware time and place perspectives, to make sure that the user meets certain criteria before receiving recommendations. As discussed before, it is important that a user is located in a different environment than their current residence, before they receive a recommendation. With enough data about users' coordinates (from all the different times they open the app), their area of residency can be mapped. This gives an indication as to where recommendations should not be given (addressing the aspect of familiarity), but also helps define the exact moment the system can consider the user as being *on the go*.

For example, if the user opens the app in a destination that is over a fixed minimum distance away from the place that the system understands as their current residence, the system continues with providing a recommendation. The *user_id* that is present in both the first and second datasets are then compared to find experience preferences for the specific user. Based on the preferences, the right experience is recommended to this user. The system essentially acknowledges that this user is *on the go* somewhere in Denmark (the risk-aware circumstance) and furthermore checks to see if experiences that fit their voting history are also present by utilizing the *content-based filtering*. If this happens to be true, a recommendation will be sent by the recommender system.

To delve further into the push notification criteria, the specific time of day is not necessary to track from the user. This is because dinner is widely eaten around 6:00 p.m. in Denmark. This means that the system does not need to get hold of any data to back this claim of sending out a recommendation at around 5:00 p.m., since it is the cultural norm for most people in Denmark to eat around 6:00 p.m. For future expansion of the system's capabilities, if more data were to be gathered about a user's eating patterns (checking in at a restaurant around 7:00 p.m., for example), there might be a reason to tweak the recommendations to meet this user's preferences. For the current iteration of the recommender system, it is only really the location that the application needs to track. This concludes the risk awareness that needs to be present in the algorithm.

To summarize, the chosen content is based on the polls that OpdagDanmark distributes and gives an indication of how users like certain experiences (their preferences, be it a price-range or selection). These preferences are linked to the *user_id*, and will be utilized to recommend similar experiences to what they have already shown an interest in. This establishes the filtering approach of the recommender system as being *content-based filtering*. As a measure of risk awareness before the recommendation is sent, the user has to be close to the experience and at the same time located a certain distance away from their area of residency. It is also important that the time of day is right for the type of experience in question.

10.3 Step 2 - Proper Data Handling

In this section, the data handling for the purpose of the recommender system will be investigated and suggestions for optimization will be given. This will point to the best conditions for collecting, structuring and tracking data to go hand in hand with the information environment and recommender system.

10.3.1 Collecting Data

OpdagDanmark already collects certain data, which can be utilized in order to back up part of the concept. As established, these include coordinates and voting history of OpdagDanmark's polls. The only place where data collection falls short is the usable information about certain experiences that will assist in building user profiles. These profiles need to be reliable enough so that it is logical that they start receiving certain recommendations. There is certainly in-depth information to be found in the content that is recommended, but this information is not plotted into any datasets that will be able to transfer into the actual user profiles and help argue for certain patterns.

As discussed in the concept walkthrough, *content-based filtering* works by assigning certain information to an existing user, from a piece of content that has already been interacted with. This slowly creates a knowledge profile for the user by building a pattern of preferences associated with certain relevant experience types. In this specific case, there needs to be sufficient data that backs up a user profile based on prior content interaction. This content interaction is miniscule for now, as currently only counted votes in polls are available. For example, after the user has expressed an interest in a certain pizza place, by voting on it in the "*Denmark's best pizza*" poll, the user profile is updated with the descriptors of the pizza place.

However, this is where a problem may occur if the experience does not have the right conditions to have its descriptors utilized for user profiling. Therefore, the many experience conditions need to be streamlined and plotted before user profiling can occur in the system. For the sake of this concept there are going to be a total of three descriptors that will be utilized to showcase the recommender system as a proof of concept. In a practical setting this could easily become many more descriptors, based on the level of precision the system is aiming for alongside the available plotted data. In the proof of concept, the experience in question is pizza places, where having three descriptors is likely sufficient to start the act of recommending content. The three descriptors are as follows:

1. **Average price for a standard 1-person pizza**
2. **Average amount of pizzas on the menu**
3. **Availability of vegan pizza choices**

The first descriptor is very basic and indicates if the place is cheap or expensive. This is a very standard dividing factor where people can have vastly different preferences based on different socioeconomic conditions. Some people will also use this to judge a place's quality of food, without actually having tried the food. Utilizing a combined average price for a pizza is therefore applicable to help specify a user's preferences, especially if the place is either extremely cheap or extremely expensive.

The second indicator considers the options available on the menu and can give a glimpse into the user's process of choosing their food option. If the user has voted on a pizza place with a large selection of pizzas, it can signify the possibility that the user likes variety. In contrast, for smaller menus, the user might have voted on a pizza place with a smaller selection because it makes it easier to choose.

The third indicator regarding vegan options on the menu is a growing matter that is worth taking into consideration. Especially since pizza places in the larger cities of Denmark are introducing vegan options. This is something to keep in mind when it comes to user profiling. If the user deliberately goes to pizza places that only serve vegan pizzas, then this is crucial knowledge for the recommender system. These three descriptors will need to be included as definable information inside the actual experiences for them to be linked to users. This means that whenever OpdagDanmark plots a pizza place into their experience map, information about price range, selection, and vegan options needs to be defined and plotted into the dataset for the experiences. These data strings can then be moved onto the user profiles whenever a user votes for a random pizza place in the polls. This concludes the examination of data collection.

10.3.2 Structuring Data

Concerning how to structure the available data and how to apply it to a user profile, it is preferable to combine the existing mixed datasets into individual datasets focused on the specified *user_id* (Figure 18 & Figure 19). The information about *voting history* and *coordinates* are currently separate data strings, and there is no system to the structure of *user_id's*. In other words, there are no actual user profiles at this point in time. Combining the relevant data for the specified recommender system (location and voting history) and dividing this data between *user_id's*, would then establish a structure for user profiles. This would give the system a starting point when recommending content. There is no need for the system to start comparing *user_id's* from multiple datasets, as all information for a single profile is located in a designated sheet. This ensures that there is no mix-up between user preferences. It also makes it faster for the system to decide upon what to recommend to each individual user.

The descriptors mentioned in the *Collecting Data* section also need to be structured in a way that fits the new data-structure. It would make sense to publicly define the experience descriptors on individual experiences inside articles and guides at OpdagDanmark. This could improve the explorative aspect of the application, regardless of the recommendation aspect. When a user signifies that they prefer a specific pizza place by voting on it in the poll, the dataset for that particular *user_id* should be updated with the new preferences too. The *user_id* profiles should then be able to store data about preferences for multiple experiences for each user.

10.3.3 Tracking Data

The user profiles that have been established in the Structuring Data section are worth tracking in a wider perspective to see overall patterns between users, and how they respond to the recommendations that they receive. It would be wise to include a rating-feature in the app that indicates if the user found their push notification helpful. With this feature, it is possible to conclude if the recommendations are worthwhile or not. The user giving ratings could be one way of accomplishing this, but as long as there is at least some sort of tracking implemented to help indicate how the recommendations perform, it will ensure an aspect of evaluation.

As of right now, there is no detailed approach to user profiling. It might be a good idea for OpdagDanmark to create accounts for their users, so that they have a log-on option on the app-domain. This could be done for the aspect of optimizing user experience and in order to track the user profiles better and, as such, be able to receive insights about the domain (although these aspects essentially go hand in hand). For example, it would be much simpler to track the hypothesis of demographic changes alongside the implementation of the push notifications that this concept puts forth (regarding the younger audience). This would be easy to realize if users were willing to state their gender and date of birth when signing up to the app.

It could also be possible to implement a way to track a particular user's interaction with certain experiences. This builds on the fact that the recommendations are, with this iteration, only based on the voting history. In some cases, this is only a single experience. Since there is a lot of potential to build upon the user profiling aspect and make it more reliable, it would make sense to implement ways for the system to know if a user has tried an experience and liked it without voting for it in the polls. If a feature that could support this profiling strategy was implemented, it would help strengthen the experience preferences of the individual users and make more in-depth recommendations. A feature that could support this indicative aspect would be something like a journaling feature that lets users log experiences that they have tried, maybe even for their friends or family to see. This is not applicable right now but might be at a later iteration of the application, after OpdagDanmark has had time to expand.

In summary, the investigation into how OpdagDanmark should collect, structure and track their data in the most optimal way is now complete, in addition to the chosen recommender system and the information environment examined in the *Structuring Data* section. Most importantly, it has been agreed that user profiles should be constructed with the provided datasets and will function as the defining factor in regard to how recommendations are argued for in the recommender system. A typical user profile dataset consists of the following strings:

- **User ID** (This information defines the specific user in question and acts as the overall name of the entire dataset)
- **Coordinates** (This is location data from app usage that is used for risk awareness. Tracking needs to occur at all times for recommendations to be viable)
- **Voting history** (Starting point for getting an idea of experience preferences)
- **Experience preferences** (Initially created with the voting history and is being utilized to argue for the relevancy in possible recommendations)

As a starting point for the chosen recommender system, this sort of dataset will be sufficient in order to argue for spontaneous push notifications *on the go*. However, this is not a particularly strong user profile strategy compared to what may be possible when OpdagDanmark is further along in their organizational maturity. This is an expected outcome, as OpdagDanmark's application is still in development and will remain so for a long time. Opportunities will arise for OpdagDanmark to tweak their data handling, which in turn can optimize profiles.

10.4 Step 3 - Formal Ontology

The OWL-file for the ontology can be found in Appendix 1.

10.4.1 Ontology Development

After having completed the analysis of the information environment, settling on a filtering approach and establishing data handling, it is now relevant to look into the implementation of a semantic layer to help logically structure the acquired data. This will be approached with Noy & McGuiness' ontology development plan.

10.4.1.1 Step 1 (Determine domain and scope of the ontology)

This step functions as a foundation for the general work and can be considered an introductory step. Essential information about the ontology will be given by answering the questions featured here:

10.4.1.1.1 What is the domain that the ontology will cover?

The overall domain of this ontology is based on OpdagDanmark's application, and will investigate the relationship between users and content, in order to argue for communicating specific recommendations with push notifications. The recommender system behind the notifications will take the user's voting patterns (experience preferences) and match them with recommendable content that match these preferences on the domain.

10.4.1.1.2 What are we going to use the ontology for?

This ontology is useful for OpdagDanmark in order to give an overview of the logical implications that lie in relevant content recommendations, based on the targeted information environment. They will be able to gain an understanding as to why a specific user should be presented with a specific recommendation, based on collected information about them. This can help signify exactly why these types of recommendations should be implemented logically. Hopefully, the users receiving recommendations will find them useful, strengthening their user experience. Furthermore, the ontology can be utilized to help establish new domains of logical user-content relationships that span multiple areas of their application.

10.4.1.1.3 What types of questions should the information in the ontology provide answers for?

In order to test the underlying structure of logical relationships between information, questions can be asked about the ontology domain. Such questions are referred to as competency questions and will help validate ways that the ontology might prove to be useful. Concerning the recommender system, there is only one competency question that needs answering. By answering this question, an argument for a specific recommendation can be built. This question is the following:

- *"Which unfamiliar pizza places match **any** specific user?"*

10.4.1.1.4 Who will use and maintain the ontology?

Fundamentally, the ontology will be used and managed by OpdagDanmark, in order to argue for recommendations, but also to scan for useful patterns in the information or to validate arguments for new functionality.

10.4.1.2 Step 2 (Consider reusing existing ontologies)

At this point in time, there are no existing ontologies based on any of OpdagDanmark's domains. There are existing templates for general recommendation, with an example of this being The Recommendation Ontology 0.3 (Ferris, Jacobson, 2010). However, this template will likely prove invaluable here, since this particular concept both acts as a proof of concept and since the fundamental work of establishing an ontology is important to showcase in this thesis. In a practical setting, templates like these could however be very useful for decreasing the workload, as long as they fit into the domain.

10.4.1.3 Step 3 (Enumerate Terms)

In this step, the terms that may be included in the ontology will be established and categorized. For this purpose, a content audit will be utilized (Hedden, 2010). The content audit works by dividing individual terms into three sections. First, the original concept name of the term, next, a variation of that specific term (to further specify), and finally, a category that can help organize the term in the larger perspective. The ontology that is developed with this iteration will remain relatively simple. However, if the ontology is expanded, and new maintainers take over its functionality, this way of categorizing might prove practically useful.

Table 1: Content Audit – OpdagDanmark.

Concepts	Variations	Categorization
Pizza place	Fast food	Content
Experience	Article/guide	Content
“Best Pizza in Denmark”	Poll	Content
Restaurant	Food experience	Content
Food stall	Food experience	Content
Take away	Food experience	Content
Animal sourced food	Includes meat, eggs, dairy	Descriptor
Vegan	Excluding animal sources	Descriptor
Pizza price	Pizza cost evaluation	Descriptor
Menu selection	Food possibilities	Descriptor
Ratings	Quality evaluation	Descriptor
Food allergy	Excluding food sources	Descriptor
Calories	Dieting options	Descriptor
Deals	Saving money	Descriptor
Voting history	Preferences	Descriptor
Hometown	Familiar environment	Place
Visitor	Unfamiliar environment	Place
City	Metropolitan	Place
Region	Country division	Place
Denmark	Country	Place

10.4.1.4 Step 4 (Define the classes and the class hierarchy)

The class hierarchy is based on a top-down approach (Figure 20). There are four overall classes present in the domain, which will help define the recommendation aspect. These are *User*, *Content*, *Descriptor,* and *Location*. The *User* class consists of each user that is interacting in the domain. Since there is no information about demographics, users do not have subclasses that divide them in terms of gender or age, but this could be implemented later. For the sake of making the ontology easy to understand, users have names. Content is supposed to be divided into all individual experiences (food, hotels, hiking trails etc.), but *PizzaPlace* has been established as a targeted area for the proof of concept, hence that will be the only experience type utilized. This class includes all the different pizza places that have been featured in OpdagDanmark's polls (as a result of the recommender system acting as a proof of concept, there are only three of them implemented in the ontology). Then there is the *Descriptor* class, which consists of information such as Pizza*FoodSource*, *PizzaMenuSelection* and *PizzaPriceRange*. These define the classes in *PizzaPlace* but will also help define users based on their voting pattern. Finally, there is the *Location* class, which essentially functions like the *Descriptor* class, by being able to argue for the location of both the users and the content. The reason that this is a class by itself, is because it can be utilized to investigate the concept of familiarity later on.

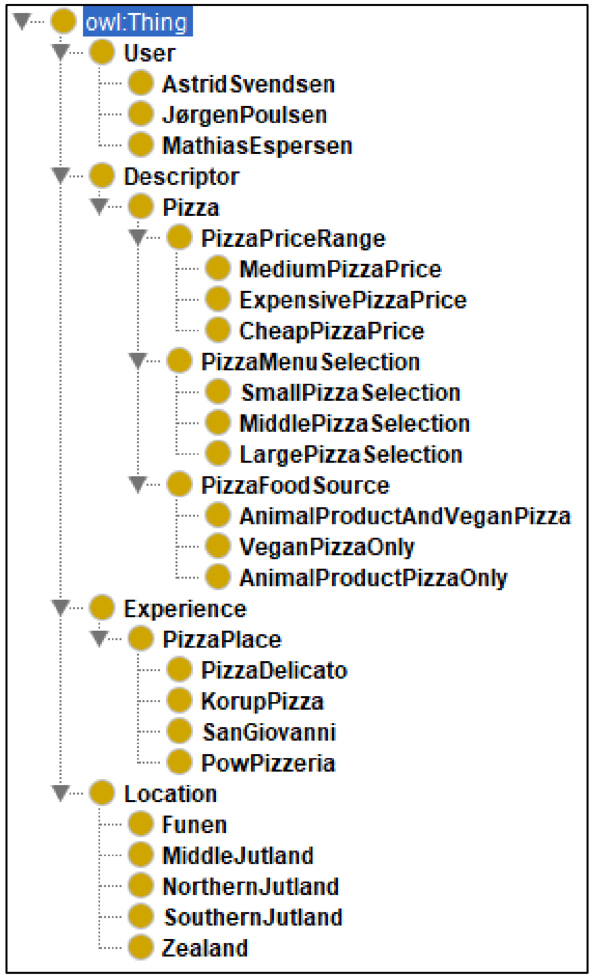

Figure 20: Taxonomy for class hierarchy – OpdagDanmark.

10.4.1.5 Step 5 (Define the properties of classes)

Now that a class hierarchy has been established, it is time to look at the properties that bind them together (Figure 21). First of all, it must be established how a *User* is voting for an *Experience*. This is done with the object property *hasVotedFor*. The *Descriptor* class is combined with *Experience* by utilizing the object property *hasProfile*. The same course of action is relevant for *Location* with the object property *hasLocation*. *hasVotedFor* is the defining object property in terms of linking user and content, thus learning about the user's preferences and where they likely live. This is accomplished by grouping all three object properties under a transitive object property *belongsTo*. With this logic, the *Descriptor* and *Location* subclasses associated with a specific *Experience* will therefore be able to define a user, which *hasVotedFor* that *Experience*.

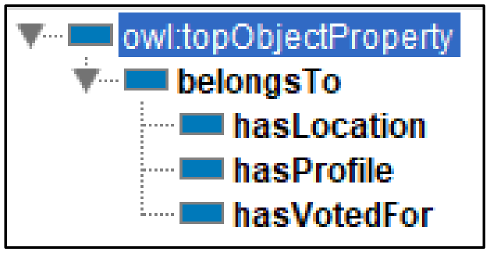

Figure 21: Hierarchical listing of object properties – OpdagDanmark.

10.4.1.6 Step 6 & 7 (Define the facets of the properties & create individuals)

There is a possibility to factor in cardinality restrictions for the three object properties, but the practical functions tied to the aspect of cardinality restrictions do not do anything valuable for the purpose of the overall ontology. For example, in practice, there is a possibility to implement a single cardinality restriction to the object property of *hasLocation*, since a pizza place cannot be in two places at once. This does not actually help OpdagDanmark decide on recommendations.

The same can be said for individuals. Right now, the purpose of the ontology exists and functions regardless of individuals, even though these could be implemented to specify the users or pizza places further. This is not to say that either cardinality restrictions or individuals could not be implemented at a later point in time, but for now, there is no solid reasoning to make the ontology more advanced than it has to be. Implementing it anyway could result in a steeper learning curve or information overload for newly assigned maintainers, which is something that should be avoided.

10.4.2 Presentation of Logic

Now that the ontology has been developed, it is time to show the core aspect of its functionality (Figure 22). The pizza restaurant SanGiovanni is being targeted. First of all, SanGiovanni is a place where they have pizzas on the menu. It is located in Northern Jutland, it has expensive pizzas, it has a small selection of pizzas, and finally, it only serves pizzas based on at least one animal source. This place has been voted for by the user Jørgen Poulsen.

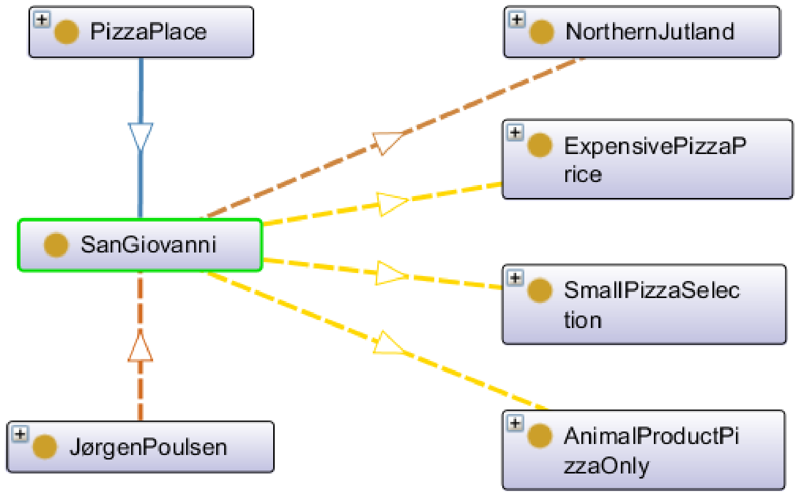

Figure 22: OntoGraf visualization of the ontology – OpdagDanmark.

With this information, it can be determined that Jørgen Poulsen is interested in Pizza. He is particularly interested in the experience of pizzas from SanGiovanni, since he voted for them in a poll that specifically tries to find the best pizza in all of Denmark. The pizzas of SanGiovanni have some distinct descriptors associated alongside a specific location in the country of Denmark. Now, with the transitive logic that is utilized to argue for *hasVotedFor* as being a defining factor in building a user profile for Jørgen Poulsen, it is now possible to take the exact descriptors that define SanGiovanni and apply them to Jørgen Poulsen as well. The logic reads:

$A = B$

$B = C$

Therefore: $A = C$

10.4.3 Answering the Competency Question

Now that the logic has been established, it is finally time to validate the ontology through the reasoner tool and see if the competency question that was established in the first step of development holds true. The competency question is:

- *"Which unfamiliar pizza places match **any** specific user?"*

To answer this question, the object property *belongsTo* will need to be utilized. The basis of finding a match to showcase the proof of concept, will be by using one descriptor. In this case, it will be regarding vegan pizza options. To do this properly, both the descriptor *VeganPizzaOnly* and *AnimalProductAndVeganPizza* have to be factored in simultaneously, as both of them indicate vegan pizza options on the menu. For the competency questions to be answered, there needs to exist a place that serves vegan pizzas outside of the familiar environment of "**any** *specific user*". *Unfamiliar environment*, as a logical concept defined by the semantic structure of the ontology, is currently being indicated with regions (*hasLocation*). Of course, the regions of Denmark are rather large, but specifying familiarity can be challenging. For the sake of keeping the ontology at a simple level, the regions function as a defining factor. It is relevant to emphasize that an *unfamiliar environment* indicates that the match cannot be the same pizza place that the user voted for already. The query for the reasoner is as follows (Figure 23):

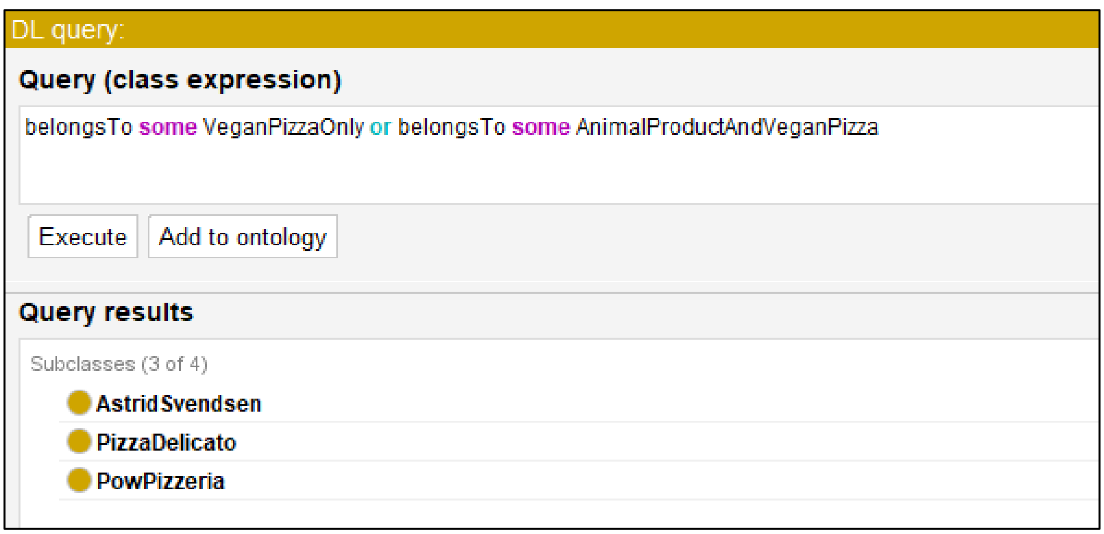

Figure 23: DL query for answering competency question – OpdagDanmark.

The result indicates that there is one user, Astrid Svendsen, and two pizza places, Pizza Delicato and Pow Pizzeria, that match the query. As stated, the user must be interested in vegan pizza, which means the reasoning process must be examined (Figure 24).

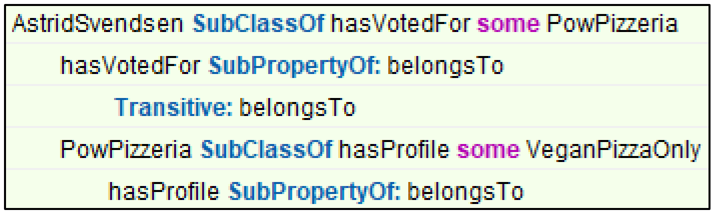

Figure 24: In-depth explanation for reasoning – OpdagDanmark.

What is depicted is transitive logic, since Astrid Svendsen has been included in the query results based on the object property *hasVotedFor*. She voted on Pow Pizzeria. Pow Pizzeria has the descriptor *VeganPizzaOnly*. With this information at hand, it makes it reasonable to assume that Astrid Svendsen is probably vegan. However, it is not completely logical to define her as vegan with the current information. She has not personally indicated that she is vegan, and the assumption is only based on a single vote. The only thing that is known for sure, is that she, to some extent, shows an interest in vegan pizzas. However, this preference still functions as an indicator for vegan pizza as being reasonable to recommend to her.

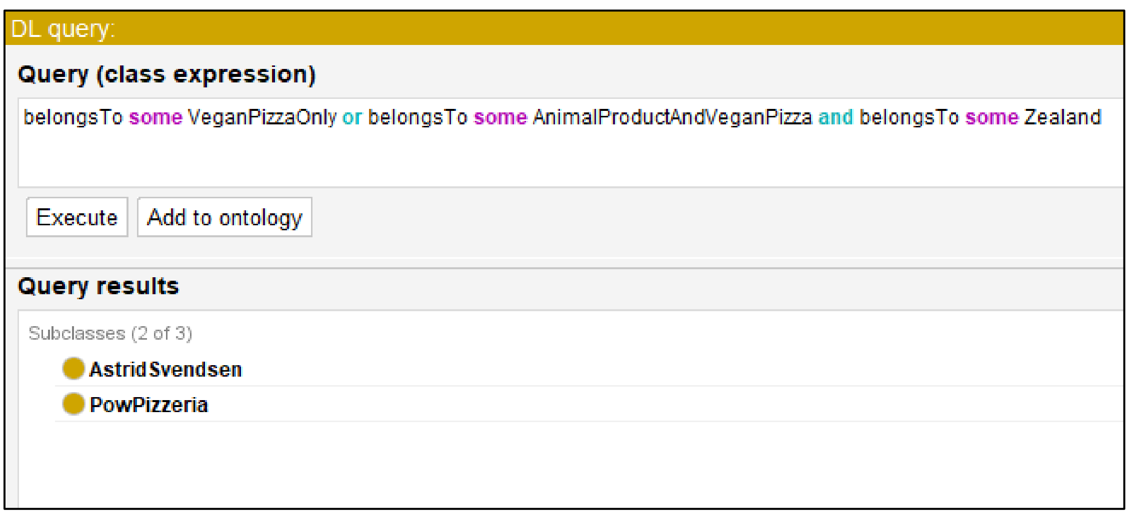

Figure 25: Expanded query to determine familiarity – OpdagDanmark.

Since location also plays a role, it must be established that Astrid Svendsen and Pizza Delicato are not placed in the same environment (Figure 25). By expanding the query to include *belongsTo* **some** *Zealand*, which is the region that Pow Pizzeria is located in (and Astrid Svendsen as well by transitive logic), it is evident that Pizza Delicato is not located in that same region. This means that Pizza Delicato can be considered an *unfamiliar experience* which meets the criteria of having vegan pizzas on the menu. Astrid Svendsen can therefore have this pizza place recommended to her. This ultimately answers the competency question by utilizing one distinct way of handling the domain of the ontology.

Of course, one descriptor is not much to base a recommendation on, but if more pieces of information were fed into the ontology, it could be possible to find very specific matches between users and content. One might consider a match based on up to ten or twenty different descriptors. As OpdagDanmark manages the ontology and decides to expand it, this might be a possibility at some point in the future. It is also important to notice that OpdagDanmark can begin to establish logically defined classes into the ontology domain. These are potentially going to be based on certain recurring descriptor classes.

For example, this could be a logically defined class concerning user preferences for specific types of menu items rooted in the details about previous food-option ordering history. However, this requires more in-depth data-collection capabilities, which are likely to be implemented into the application at a later point in conjunction with the introduction of user profiles, gamification elements etc. It is important to state, that these options are not a part of this concept but are possible to implement on top of the existing semantic layer if OpdagDanmark so chooses.

10.4.4 Interaction Graph

In order to simplify the core idea of the ontology, and simultaneously showcase how it merges with the chosen filtering approach, an interaction graph has been created (Figure 26). This visualization shows how the user, who votes for an experience, receives the characteristics of the experience as defining (profiling) information, which in turn is being used as an argument for recommending a similar experience. The crucial aspect of being recommended experiences, which are located in an unfamiliar environment, is also showcased here through the different regions (listed for each experience).

The outer ring around the user signifies that the targeted profile, which the recommendations are based on, could include more users simultaneously. This is an example of utilizing the logically defined classes mentioned in the *Presentation of Logic* section. This *class* could be named *CheapVeganPizzaPreference* (defined by the experience descriptors that signify a preference for *Cheap* and *Vegan*) and could possibly have a number of users (subclasses) attached to it in the formal ontology. This is only relevant to define when the general user base on the domain is large enough.

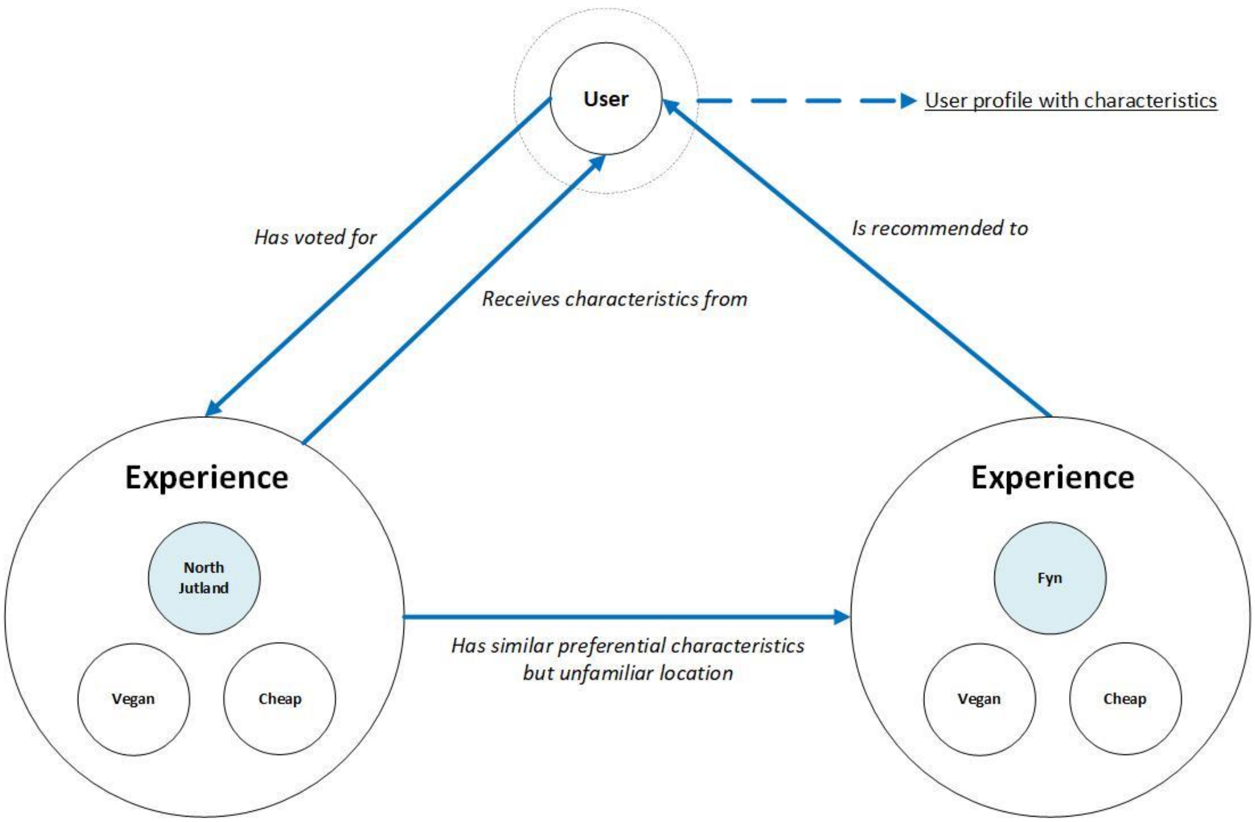

Figure 26: Interaction graph of recommender system – OpdagDanmark.

10.4.5 Ontology Evaluation

Now it is possible to evaluate the current iteration of the ontology. It is obvious that the current iteration is just the beginning of what is possible, since it only focuses on one distinct experience type. However, with the current proof of concept, it is shown that a semantic layer is indeed able to deliver value in practice and can even be utilized in a number of different ways. The vegan example was one distinct way of using the knowledge domain, but with different descriptors it might be a completely different approach and a different way of interacting with the ontology altogether. This encourages the person maintaining the ontology to think creatively and come up with reliably built and detailed arguments for what to recommend and to whom, as this human aspect could drastically increase the likelihood of interaction.

The nature of push notifications may be a risky element to implement into a user experience. As such, it is good practice to only use them in instances where value is almost certain to be created for the user. However, push notifications should still not be automated too liberally, as this could easily result in a lot of people uninstalling the app altogether in frustration. Automated machine learning may prove to be too unreliable for this activity, as the final recommendations can sometimes be illogically accounted for. This further proves the value of a semantic layer, as a human aspect is valuable in risky communication (such as push notifications). Since OpdagDanmark is a small startup, there is simply no room to make too many mistakes. Transparent and logical arguments might therefore be the best approach to strengthen their user experience. This summarizes the reasoning for what content to recommend to certain users. The next step will look more closely into risk awareness in terms of *time and place*.

10.5 Step 4 - Persuasion

In terms of this particular concept, it is most relevant to attach persuasion to the existing risk awareness that is associated with push notifications. The persuasion that has previously been discussed in the terms of *time and place* (the Kairos-moment) will function as a way to minimize the inevitable risk associated with the recommendations. Push notifications are, by their very nature, intrusive, and the premise for this concept to create value for OpdagDanmark will require the system to deliver push notifications. Therefore, the persuasion should not be seen as a way of enhancing the recommender system per se, but more as a tool to further minimize the risk of users being irritated by the notifications, making them more accurate by taking *time and place* into consideration.

Kairos is utilized by first examining the possibilities attached to the aspect of time. As was briefly discussed in *Step 1*, there is a good chance that people who are looking to eat pizza will do so around culturally defined dinner times. This means that in order to persuade the user to try a specific pizza place, the notification must be sent before the user is expected to eat. Although not too long before, as the sporadic nature of the targeted user might not be persuaded by the notification if sent long before they sense an obligation to start looking for dinner possibilities.

This also brings in the subjective aspect of time, which begs the question; is the user hungry? It would be perfect timing to send the notification if it is guaranteed that the user is hungry, but this information is practically impossible to predict with data from the application alone. Maybe if the user was actively looking through restaurants in the app, then it could be safe to assume that providing a recommendation would be appropriate. However, this would not account for the aspect of *serendipity* (Smetsers, 2013), which is also part of making the notifications worthwhile. It is presumed to be more reliable to go with the typical patterns for eating dinner in Denmark (between 6:00 and 7:00 p.m.) and try to persuade the user with a notification that is permitted by the system to be sent between 4:00 and 5:00 p.m., for instance.

The *time* aspect of the risk awareness has to be partnered up with the aspect of *place* as well, so that these two aspects can be matched with each other before the system sends the recommendation. *Place* is more interesting to investigate in this practical case, as there is data for coordinates associated with user profiles. This data can be utilized to ensure that the user is in close range to a pizza place that meets the requirements for recommendation. As of right now, OpdagDanmark only tracks coordinate data associated with users opening their application. If the recommender system is to work as intended, it will need to track continuously.

The current location of the user also needs to be considered. If the user is walking, they might not find a recommendation relevant, if it is for a place that is 4 or 5 kilometers away from their current position. This, of course, depends on the user's mobility context (bicycle, car, bus etc.). The subjective aspect of *place* is therefore linked to how the user is moving around in their current location. This could probably be tracked in terms of speed. Low speed implies walking, medium implies cycling and fast speed implies driving. The higher the speed, the higher the possible distance gap between the user and the recommendation.

Another factor to consider could be urbanization. For example, if the user is in the middle of Copenhagen, then 2 kilometers may seem very far away. It is worth considering that there also may be many eating options much closer in proximity to the user. The same distance is perceived as dramatically less problematic, if the user is traveling on a straight road in a non-populated area, where other similar food-options are more scarce.

A subjective aspect of *place*, which was accounted for in the ontology, is the sense of familiarity. If the system knows that the user is in a familiar environment, then it should not recommend anything. Right now, this data is based on the *hasVotedFor* object property in the ontology, which essentially functions as a filter. This is not completely logical, since it only takes voting into account. If the tracking of location is permitted continuously, this filtering aspect could then be changed to be based on this data alone. This would create a more precise measure of familiarity/unfamiliarity for all users in the database. The downside is that this practice can, to some extent, be considered a privacy-breach. A more in-depth discussion of the implications associated with this will be reviewed in the *Discussion* chapter.

Now that the circumstances for *time and place* have been examined, it is appropriate to examine the actual push notification. It is a simple means of communication, as a push notification is not very customizable and does not contain much information. Instead, its function is to trigger the user to access the pizza place's article on OpdagDanmark's application to further persuade them to seek out the experience. The push notification needs a rhetorical approach that is going to persuade, and at the same time seem genuine, like it is being communicated manually by OpdagDanmark (which it to an extent is with the semantic aspect). It also needs a great visual hook to cater to the user's hunger, and make the person crave pizza.

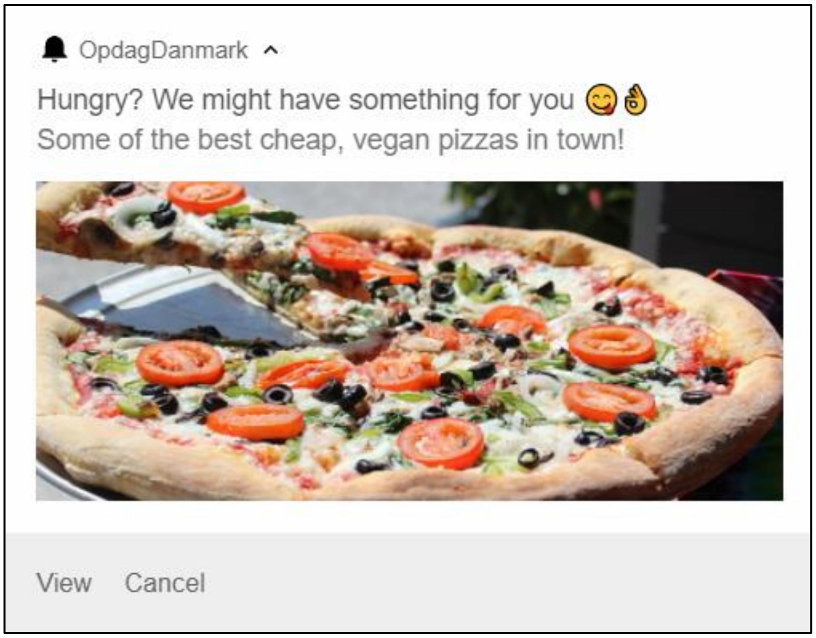

Figure 27: Example final recommendation – OpdagDanmark.

What is presented above (Figure 27) is an example of a push notification for a pizza place that has the *cheap* and *vegan* descriptors attached in the ontology. Notice that these descriptors are utilized as selling points. The emojis reflect OpdagDanmark's own means of communication, and at the same time make the notification seem warm and welcoming. The "*We might have something for you*" signals a proposition; OpdagDanmark is giving something (a piece of information) to the user. This can help signal that the notification is part of OpdagDanmark's overall product, and hopefully seem genuine to the user (instead of merely giving connotations of general advertising). Lastly there is a picture of a vegan pizza attached, helping the user visualize the food in front of them already.

10.6 Evaluation

Now that the recommender system has been properly established for the practical case, it is time to look at the overall product. First, a walkthrough of the user's journey in conjunction with the recommender system. The user engages with the app, which establishes a dataset for the exact ID (profile). When this profile has voted for an experience (through the polls in the app) the recommender system can now index the user's ID as available for recommendations. With the help of the ontology, the system finds the most recommendable experiences (similar, yet unfamiliar) in conjunction with the experience that was voted for by the user. These recommendable experiences will then be linked to the user and primed for an eventual push notification.

However, this will only happen if the user meets certain criteria (risk awareness). In the case of pizza places, there are three sets of criteria. The first one is that the user must be within a certain number of kilometers away from the pizza place (based on mobility context, which the system detects by measuring how fast the user is moving). The second one is that the time of day is between 4:00-5:00 p.m. (the recommendation is simply caught by the system's risk-aware filter if outside of this timeframe). Finally, the user must be located in an unfamiliar environment (reducing the chance of them being aware of the recommendation already). The push notification will then be sent to the user.

Evaluating the system and its development with OpdagDanmark, they have been happy with the results. As mentioned earlier, they are planning to expand upon the existing application, so there are possibilities to extend the targeted information environment of the ontology. The data that was initially given from OpdagDanmark has been sparse in terms of utilization possibilities, but this is not a problem, since the proof of concept-approach puts emphasis on the possibility of expanding on the initial information environment. Things like logging user-profiles, continuously tracking locations, and more practical and detailed information for every available experience on the map, will inevitably create an information environment that can argue for more and more reliable in-depth recommendations. The starting point that has been established with this case helps to pave the way for OpdagDanmark's semantic utilization of their data, but it is nowhere near its intended capacity.

11 Practical Case For NNIT

11.1 Initial Meeting

To initiate the collaboration with NNIT (specifically the CoMa-department) an initial meeting was arranged. The attendees included the consults, and three employees from the CoMa-department: Lars B. Petersen, the *Vice President* and *Head of Communication and Marketing* in NNIT, Birgitte Barslund, a *Marketing Manager*, and Peter Nimand Jansen, a *Marketing Consultant*. These three individuals are all responsible, to some degree, for nnit.com and some of the content that is present on the site. After they had been introduced to the concept, the immediate goal was to establish the purpose of the recommender system in the case of NNIT (specifically for the CoMa-department). This would, at the same time, assist in specifying the exact domain for the recommender system.

At the meeting it was explained that there are currently no complex automation systems that deal with the recommendation of content on nnit.com. At the time of the initiation of the practical case, only a few recommendations of content are made on the website. These recommendations are created manually and are based on the content manager's intuition. As a result of the lacking aspects of content recommendations, it was decided that the CoMa-department would be interested in a complex solution that provides content recommendations for the visitors on nnit.com. The solution that is to be provided, should be a design of this system for a specific subdomain. This design would act as a proof of concept and should therefore have the capabilities to be expanded to a larger domain or multiple subdomains. As such, the domain for the recommender system was defined as the area of *Cybersecurity* on nnit.com. At the time of initiating the practical case, even though all areas on nnit.com are dynamic, some are expected to change more than others. *Cybersecurity* is currently one of the more stable domains.

With the specified purpose and domain for the coming recommender system being defined, further collaboration between the consultants and the CoMa-department were discussed. It was decided that Peter Nimand Jansen would be domain-responsible for the project, as he is responsible for *Cybersecurity* in the CoMa-department at the time of the case's initiation. He will provide the consultants with the required data, in order to investigate the targeted domain with the Information Ecology Model. This data is to be provided in the form of extensive lists of users engaging with the content on *Cybersecurity*. When dealing with personal data about visitors, confidentiality protocols must be upheld. This will be no problem however, as all data received from the cases are anonymized to ensure confidentiality.

11.2 Step 1 - Business Analysis

11.2.1 Information Ecology Model

As has been mentioned in the theoretical walkthrough of the concept of the thesis, the Information Ecology Model will be utilized to provide a holistic mapping of the defined domain, leading to how NNIT should implement a recommender system in practice.

11.2.1.1 Context

The domain of *Cybersecurity* on nnit.com (Figure 28) has different purposes. For the general public, the domain attempts to inform interested visitors about the general purpose and value of cybersecurity, with an aim to spark a deeper interest in the topic. For people in the industry, the domain attempts to provide detailed knowledge about specific principles and capabilities of cybersecurity and establish NNIT as thought-leaders on the subject. Finally, the domain attempts to convince potential customers of NNIT's services in cybersecurity to procure their services, or to push them further along in their journey to become buyers.

Figure 28: Frontpage of Cybersecurity area on NNIT with the solutions menu – NNIT.

As has been established, the CoMa-department is interested in a competent solution that can provide content-recommendations to the visitors of nnit.com, specifically for the domain of *Cybersecurity* in this case. The domain contains a large, consistent group of visitors, as well as a steady stream of distinct content. As the domain is placed on the platform of a website, certain aspects of the recommender system must be considered and upheld in order to accommodate NNIT's requests. The recommendations themselves must not stray from the overall design, identity, or purpose that NNIT is attempting to signal with their website. The CoMa-department is open to creative ideas about how the recommendations are presented to the visitor, as long as nnit.com's structural and visual identity is accounted for. This means that, depending on the functionality of the soon-to-be defined recommender system, the visual presentation of the recommendations themselves can be visualized in different ways, as long as they adhere to the identity of nnit.com. Defining the specific visual presentation of the recommendations is an ongoing collaborative process with the CoMa-department throughout the case.

11.2.1.2 Content

The content that is presented on *Cybersecurity* is varied and consists of the following distinct content types:

- Podcasts
- Videos
- Webinars (and webinar recordings)

- Health checks (assessment of company)
- Conferences (and conference recordings)
- Articles (security insights)
- Whitepapers
- Brochures
- Briefs

The current general strategy, regarding content on the domain, is to present the user with a lot of different content. As such, visitors are most often presented with multiple options of different types of content, placed statically on the website. The content and types of content that are the most substantial, are presented to the visitor more often than other lesser content. This is a result of content creators and managers of the domain attempting to push and prioritize some content over other content. Though this strategy may be effective for presenting visitors with content that the CoMa-department deems the most important, it neglects any personalized recommendation.

As it has been established in *context*, the domain has three different purposes. These different purposes also fit with the content that is being produced in the domain. The content on the domain is not only varied in type, but also in purpose and level of technicality. Utilizing the example of *whitepapers*, the specific *whitepaper* that is presented on the main page of the *Cybersecurity*-area on NNIT.com (Figure 29), provides broad knowledge of the topic in general. However, delving deeper into the domain, visitors may be presented with other *whitepapers* such as the *whitepaper* regarding "Identity and Access Management" (Figure 30). The information in this *whitepaper* is much more technical and complex, delving into specifics and challenges in the subarea. These *whitepapers* are prime examples of how content of the same type can be vastly different in their purpose and technicality. The last requirement for showcasing content on the domain is for the purpose of attracting and/or persuading potential buyers (Figure 31).

Figure 29: Whitepaper cover for IAM – NNIT.

Figure 30: Whitepaper cover for NNIT Cybersecurity – NNIT.

CYBERSECURITY

VR Cybersecurity Training

How quickly can your employees spot potential security breaches?

Cybersecurity is crucial for businesses to survive in a world dominated by cybercrime. In response, organizations are adopting numerous tools to combat the attacks, but few know that **up to 90% of cybersecurity breaches are caused by human errors**. Therefore, an embedded culture of cybersecurity awareness among employees is the most important measure against cybersecurity threats.

NNIT Cybersecurity Training in Virtual Reality is a new way of providing awareness to employees. In the experience, you get a 360° picture of your office environment where 5 potential security breaches have been placed. The challenge for the user is then to find the breaches the fastest. Gamification has proven to be a vital element in order to increase knowledge retention and strengthen training effectiveness.

Figure 31: Article on training in new cybersecurity advancement – NNIT.

11.2.1.3 Users

As has been mentioned, the domain has a large number of visitors (users). Generally, for the domain of *Cybersecurity* on nnit.com, all people between 18 to 50 years of age, who have an interest in cybersecurity, can be considered a person of interest. This broad definition makes categorizing visitors challenging. The CoMa-department attempts to combat this by primarily targeting visitors who are connected to companies with 300+ employees. However, this can only be accomplished if the visitor provides their data to NNIT. Therefore, visitors who do not provide their data are rarely targeted. As such, a new categorization of visitors is required for designing the recommender system. Regarding the sections of *context* and *content*, there is an obvious similarity in the pattern between the purposes of the domain, and the purposes of the content. The pattern that is seen between the two sections also suggests a potential categorization of the visitors (users) of the domain. In order to match the purposes of the domain and content, the users should be categorized based on their purpose for visiting. This way visitors can be divided into three groups where the purposes for visiting matches with the purposes of the domain and the purposes of the content. The three user groups are established as follows:

Top-level - This group is interested to learn about the general value of cybersecurity for their company through its general aspects and purposes. They have little to no technical knowledge about the field.

Mid-level - This group is experienced in the field of cybersecurity and has an interest in keeping up with new advancements like tools, technical considerations, and the applicability of the technology. They are very knowledgeable about the technicality of the field.

Bottom-level - This group primarily consists of people who may be potential buyers or at least decision makers. These visitors are interested in how NNIT's services fit their own organization and strategies. They are also interested in knowledge regarding how to approach and train employees in cybersecurity.

With the three established user groups, the types of users themselves must be specified. Who are the types of people that are to reside in these three categories? The website, nnit.com, already has established data-collection channels. Through Google Analytics and NNIT's CRM-system, the website is constantly gathering data on users and their behavior. Through these channels, NNIT can provide the consultants with datasets. It is these datasets that are examined and used to gain an understanding of the type of user that interacts with the content on *Cybersecurity*.

The received datasets contain data on individuals who have engaged with a piece of content, in this case a *webinar*. From this *webinar*, NNIT has permission to gather data from attending users. This data gives an overview of the different types of individuals who interact with that kind of content. Of course, as these datasets do not specifically deal with visitors on nnit.com, it becomes challenging to know if the types of users who attend the *webinars* are the same types of people who visit the *Cybersecurity* domain. However, it should be noted that these *webinars* are included amongst content that is presented in the domain on nnit.com. As such, it is reasonable to expect that the individuals, who are engaging with the content that is present on the domain, also visit, or have the potential to visit, the domain itself. The domain expert for *Cybersecurity*, Peter Nimand Jansen, also states that the types of individuals that are present in the CRM-datasets regarding events (in this case the webinar), are also the types that are visitors of the domain on the website. It must also be stated that a lot of the users, who attended the webinar, have signed up for it, through the online forms (webforms) on the domain (Figure 32).

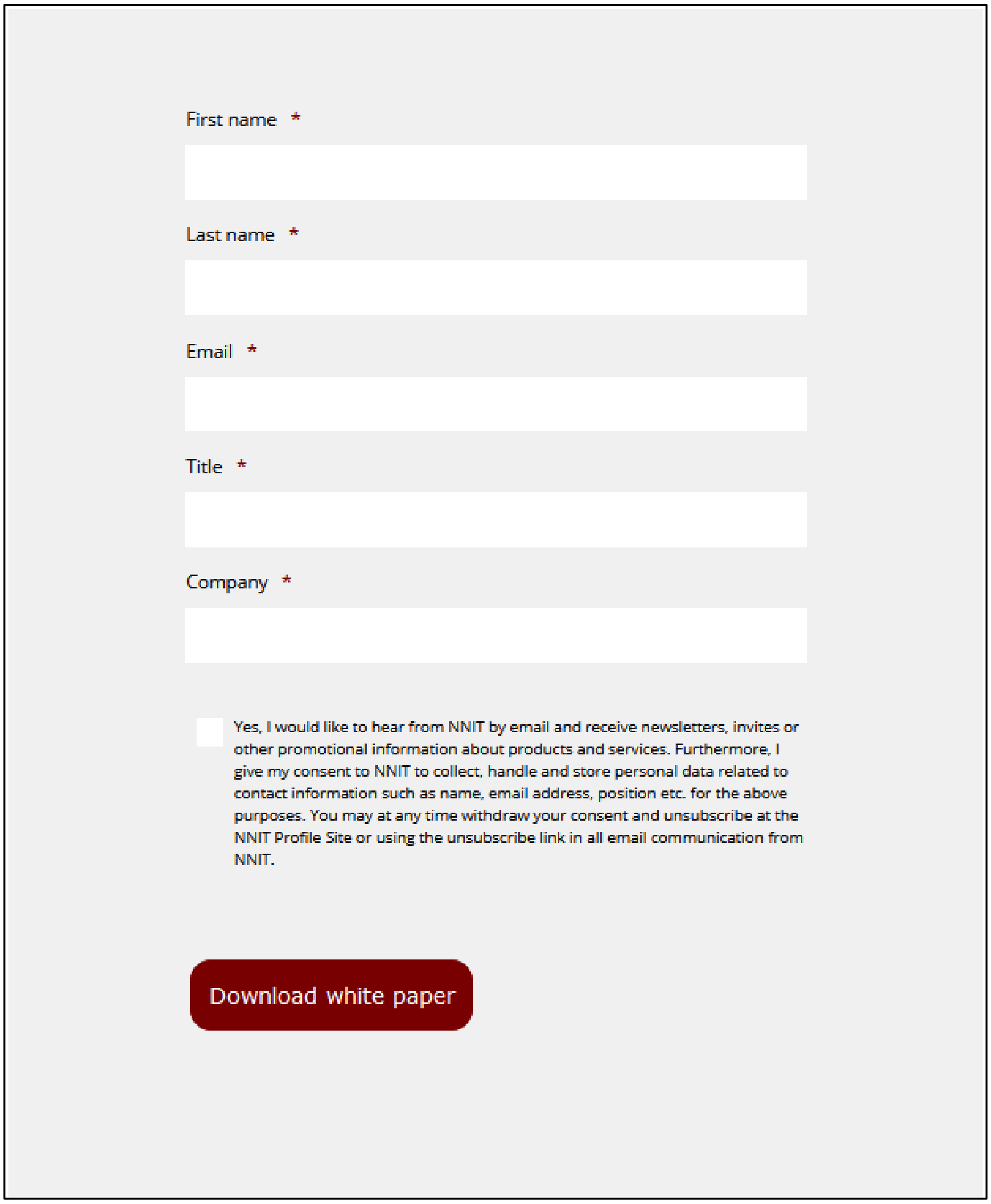

Figure 32: The webform-element on nnit.com – NNIT.

In the datasets it is even documented that some of the users have already downloaded a *whitepaper* before attending the *webinar*. With the example of the webinar's users being part of the targeted audience of the domain, the received datasets (Figure 33) can be trusted as a showcase of the types of users for the *Cybersecurity* domain.

Dept.	Job Title	Involvement	Industry	Function
Platform	Platform Owner	Rådgiver/konsulent	Offentlig administration	Konsulent/Rådgiver
It	System Administrator	Er ikke involveret	Andet:	Administrator
Information Security	Information Security Advisor	Rådgiver/konsulent	Andet:	Anden IT-stilling
It Security	Security Architect	Er ikke involveret	Sundhedssektoren	Analytiker
Global Information Security Management	Senior IT Security Architect	Delvisbeslutningstager	Andet:	Systemanalytiker eller systemdesigner
	Solution Architect	Er ikke involveret	Offentlig administration	IT-arkitekt
	IT Ingeniør	Rådgiver/konsulent	Andet:	IT konsulent
Koncern It	Firewall- OG Netværks-Specialist	Delvisbeslutningstager	IT - service	Netværksadministrator
	IT Security Specialist	Rådgiver/konsulent	Andet:	Anden IT-stilling
	IT-Sikkerhedsspecialist	Rådgiver/konsulent	Offentlig administration	Anden IT-stilling
	Beredskabskonsulent	Er ikke involveret	Offentlig administration	Konsulent/Rådgiver
Cyber- Og Informationssikkerhed	Security Coordinator	Er ikke involveret	Sundhedssektoren	Sikkerhedschef
Cfcs	Special Konsulent	Rådgiver/konsulent	Telekommunikation	Konsulent/Rådgiver
	Chefkonsulent	Rådgiver/konsulent	Energi og forsyning	IT konsulent
	SR. Compliance Engineer	Rådgiver/konsulent	Industri	Juridisk rådgiver
Information Security	ISO	Rådgiver/konsulent	Forretningsservice	Sikkerhedschef
Ito	Systemprogrammør	Er ikke involveret	IT - service	Applikations- el. systemprogrammør
It Management	IT Security Manager	Delvisbeslutningstager	Forsikring	Sikkerhedschef
Security	Ciso	Rådgiver/konsulent	Finans	Sikkerhedschef
It	System Administration	Delvisbeslutningstager	Medier/Forlag/TV	Systemadministrator

Figure 33: Extract of dataset – NNIT.

As some of the data that resides in the provided datasets is highly confidential, complete anonymity is required for the use of the data. As such, only information regarding a user's department, job title, involvement, industry, and function is utilized. The specific webinars themselves are not to be disclosed either. However, the available data still provides a relevant overview of the professional type of individual for the domain, allowing the types of individuals in each user group to be defined as follows:

Types of individuals in the top-level user group:

- They can be from any industry as long as it, in some shape or form, has a need for, or will have a need for, the capabilities of cybersecurity.
- People with non-technical IT job roles in companies that are not necessarily related to cybersecurity.
- They can have any function.
- They are not decision-makers regarding choices of cybersecurity in their organization.
- They have no, or very little, knowledge in the field of cybersecurity.

Types of individuals in the mid-level user group:

- IT is an important aspect of their industry, and there is a distinct need for cybersecurity within it. Examples include Public Administration, Healthcare, Finance, Telecommunications, and Transport and Logistics.
- People with technical job roles in companies that are related to cybersecurity. Examples include IT-Security Specialist, Security Coordinator, Software Manager, Risk- and Identity Manager, Solution Architect, or e-Commerce Manager.
- Examples of their functions include Analyst, Consultant, Programmer, Manager, or Engineer.
- They are not necessarily decision-makers regarding choices of cybersecurity, but they may have some technical authority.
- They have some or a lot of technical knowledge in the area of cybersecurity.

Types of individuals in the bottom-level user group:

- Some aspects of their industry are involved with IT and have, or will have, a need for cybersecurity. Examples are much the same as the mid-level user group and include Public Administration, Healthcare, Finance, Telecommunications, and Transport and Logistics.
- People with administrative and leading job roles. Examples include IT-Executive, Board of Directors Member, Chief Technology Officer, Chief Information Officer, and Head of Infrastructure Services.
- Examples of their functions include IT-Leader, Security Leader, Owner, and Administrator.
- They are decision-makers regarding cybersecurity in their company/department.
- They have average or advanced technical expertise with cybersecurity.

11.2.2 Evaluation

While NNIT is able to collect detailed personal data about the people that engage with their content in the *Cybersecurity* domain, it is only possible if the visitor provides their personal data, like when filling out a webform. A problem arises when realizing that content, like *videos* and *articles*, does not require the visitor to provide their information in order to engage with it. While visitors that are to be categorized in the *Mid-level* and *Bottom-level* user groups are expected to engage with content that requires them to provide their personal data, visitors in the *Top-level* user group are not. The individuals from the *Top-level* are not looking for specific or advanced knowledge about cybersecurity, which is often presented through content that requires the submission of a webform. Therefore, some visitors in the *Top-level*, are expected to only ever engage with content like videos and articles. As such, an individual visitor in the *Top-level* may not provide their detailed personal data that they would otherwise provide through the webforms. As one of the purposes of the *Cybersecurity* domain targets this specific type of user, developing a system that solely categorizes its users based on detailed personal data would neglect the *Top-level* visitors. Instead, by dividing users into three groups that categorizes based on the user's assumed purpose for visiting the site, no user will be neglected in the system, as it is not the individual's detailed personal data that is the catalyst for categorization.

When discussing the three established user groups, the prioritization of each group should also be considered. As mentioned, the CoMa-department has prioritized users that function as decision-makers and are connected to a company of over 300+ employees. This begs the question: Is one user group more important than the others? From a business perspective, one could assume that the established purpose of informing the general public and sparking interest in the topic of cybersecurity (*Top-level*) is not as important as the purpose of persuading potential buyers (*Bottom-level*). However, while persuading potential buyers (*Bottom-level*) is important, a part of persuading these B2B-customers is to showcase NNIT as the most knowledgeable option on the market, which fits the purposes of *Top-level* and *Mid-level*. As such, none of the three domain purposes (and therefore also user groups), should be prioritized more than others in the final recommender system.

To summarize the overall findings and ideas of the analysis with the Information Ecology Model, there are three specific purposes for the domain of *Cybersecurity*:

- Spark the general public's interest in the purpose and value of cybersecurity.
- Establish NNIT as experts and thought-leaders with deep knowledge of the field.
- Convince potential buyers to choose NNIT for providing services related to cybersecurity.

There are also many different types of content in the domain. The different content of the domain matches at least one of the three established purposes of the *Cybersecurity* domain. The visitors on the domain are part of one very broad user group. However, the visitors can be categorized into three major user groups that fit the three purposes of the domain. The reason for this being that visitors who are put in one specific user group, can then be presented with content that is related to the corresponding purpose of the domain. As such, these user groups can be defined as different *presentation domains,* where each domain contains different content that can be recommended to the related users.

11.2.3 Recommender System

The first action of defining the recommender system is to establish how the recommendations themselves should materialize. Currently, nnit.com is constructed with different elements. These elements are the different building blocks that create web pages. The aforementioned webforms are an example of one such element. The way the website is constructed is important to consider when deciding how the recommendations should be presented on the website. In order to fit the recommendations with the current identity, construction and design of nnit.com, it was decided that the recommendations should be presented through a custom element and placed at the bottom of the webpages. This element should showcase recommendations through a window-like structure, presenting three or more recommendations. This way of presenting the recommendations will also assist the exploratory nature of the webpages, presenting a visitor with a new option of content, when they have finished with the content they were interacting with.

11.2.3.1 Deciding on a Filtering Approach

When deciding on the optimal approach for filtering in the recommender system, the developed analysis of the domain must be taken into account. As has been mentioned previously, the visitor is not required to give their personal data to interact with certain content. With the system not being able to filter based on detailed personal data, a solution would be to filter based on the user's immediate actions on the site. This points the recommender system towards a *session-based filtering approach*. With a *session-based* recommender system, focus is entirely on the visitor's actions in a specific session. This removes the issue of having a need for immediate personal data about the visitor. However, session-based *filtering* alone does not provide a system that is complex enough for this case. The filtering approach must also take the three user groups (*presentation domains*) into account in order to be able to recommend relevant content. In order to accomplish this, *collaborative filtering* (in combination with *popularity-filtering*) must be utilized in the recommender system as well.

By combining *collaborative filtering* and *popularity filtering* with *session-based filtering*, the system should be able to obtain knowledge of a given visitor in a specific session. The *session-based filtering* is able to categorize the user into one of the three established user groups (*presentation domains*). This filtering is based on the initial interaction between the visitor and a piece of content. The *presentation domains* then become the "domain" in which a user is recommended content that a similar user has engaged with (as both users reside in the same *presentation domain*). This is how *collaborative filtering* is utilized. *Popularity filtering* will then assist in narrowing down the recommendable content, as a lot of content is related to a single *presentation domain*. This is accomplished over time by rating the different content in terms of the interest that is shown for it by visitors. As such, there will be content that is more popular than other content in the *presentation domain*. This will assist the recommender system in deciding which piece of content to recommend. Utilizing *session-based*, *collaborative*, and *popularity filtering* entails the recommender system utilizes a *hybrid filtering* approach, consisting of the aforementioned filtering approaches.

As mentioned previously, with regard to the *session-based filtering*, the system needs to be able to filter users based on their initial interactions with content in a session. This means that the system needs to determine when a visitor can be considered as having interacted enough with a given piece of content to be placed in a *presentation domain*. As with any website, some visitors may visit the website by mistake, or fail to find the information they are searching for, resulting in them leaving the site without any interactions/conversions occurring. Scenarios like these need to be ignored by the recommender system, so that it only filters relevant visitors. As such, criteria for when the *session-based filtering* should start filtering needs to be established. These criteria also need to ensure that the recommender system will filter the visitor into the correct presentation domain. This needs to happen only after the visitor has been engaged with a piece of content for a certain amount of time, but also before the visitor is finished with the content, as there is a risk of the visitor leaving the page without being recommended content in time. In order to define the specific amount of time before a user is filtered, the length of *articles* should be taken into account. However, this example only works with some of the content that is present in the domain. With other examples, where a user has to make a conversion (download a *whitepaper* or sign up for a *webinar*), the user should only be categorized into the correlating presentation domain when the conversion has taken place.

The criteria for making *session-based filtering* work, is also an option for establishing how the *popularity filtering* should work. It can be argued that the more frequently a given piece of content has achieved its criteria (been interacted with), the more popular it is. As such, the criteria become a descriptor for the *popularity filtering*. Through data collection of the number of achieved criteria for a given piece of content, the system will be able to define the most popular piece of content.

11.2.3.2 Circumstances of the Recommender System

Like with any recommender system, the circumstances and risk awareness of the actual recommendations need to be defined for the system. As has been established, the recommendations are to be presented on nnit.com, specifically in the domain of *Cybersecurity*. “When” the recommendations should visually occur was briefly discussed as being when a visitor has achieved the established criteria of being categorized into one of the three *presentation domains*. What remains to be defined is the number of recommendations to be given. As there are many different types of content, one could be tempted to present the visitor with everything related to their given *presentation domain*.

At first, this would also seem like a good idea, as NNIT's prior strategy has been to present the user with as much relevant content as possible. However, the recommender system must act differently and focus a specific selection of content toward the visitor. Even without the recommendations, users are presented with a lot of different content to choose from. Adding an excess number of new recommendations would discourage the visitor by overstimulation and lack of ability to focus on all the content simultaneously.

Besides these circumstances of the recommender system the persuasion of the recommendations should also be considered. For this case, it is expected that B.J. Fogg's Behavior Model (Fogg, 2009) will be invaluable. As the model delves into the *motivation*, *ability,* and *prompts* of the users, it seems relevant to introduce it here, considering the purposes of the *Cybersecurity* domain. As the established filtering ensures distinct content for each of the *presentation domains*, it becomes invaluable to understand whether the visitors that are filtered into the *presentation domains* are motivated to engage with the recommendations. Utilizing the B.J. Fogg's Behavior Model may provide insights into how recommendations should be presented.

11.3 Step 2 - Proper Data Handling

Step two of the practical case for NNIT delves into the data handling practices that are currently present in the domain. This section will also discuss data handling practices that are not currently present in the domain but may be needed for the purpose of improving or expanding the final recommender system.

11.3.1 Collecting Data

As has briefly been established, NNIT currently collects data for nnit.com through different channels (Google Analytics, CRM-system). They provide knowledge of the visitors' detailed personal information. However, the filtering of users is accomplished through the predefined *presentation domains*. As long as the expected visitors of a given *presentation domain* are based on real data, such as the received datasets, further collection of detailed personal information about visitors becomes obsolete for the current iteration of the system. The initial data that is required for the system to know, is whether a visitor of a session has achieved the criteria of a given piece of content, and which *presentation domain* the interacted piece of content is related to.

The three predefined groups (*presentation domains*) follow the purposes of the domain and its content, and as such are deemed sufficient for this iteration of the recommender system. However, three predefined groups (*presentation domains*) may still be very broad and non-specific, considering that the overall targeted audience is visitors between 18 to 50 years of age, with an interest in cybersecurity. As such, a clear future expansion of the recommender system would be to define more *presentation domains* from which to provide more personalized recommendations.

The currently defined iteration of the recommender system also acts as a proof of concept, in the sense that the chosen domain is very specific. To increase the practical value for NNIT (the CoMa-department) the concept has to be scalable to other parts of nnit.com. If the concept is to go beyond the boundaries of *Cybersecurity* on nnit.com, data regarding the other areas' purposes, content and users would be required. While it would make sense that some of the other areas of nnit.com have similar purposes to that of *Cybersecurity*, it may not necessarily be the case. There may be many different needs and purposes depending on the area on nnit.com.

It is also important to note that the current iteration of the recommender system focuses on what is already present in the domain, which are the three major purposes. As mentioned, users are not required to provide their detailed personal information, in the case of interacting with videos and articles, for example. However, if nnit.com at some point changes the user experience into something that requires the visitor to provide their detailed personal information, then the filtering approach of the recommender system might need to change as well. A hypothetical example of this would be if nnit.com includes user profiles (logins), or visitor questionnaires. This would provide a more stable and reliable stream of detailed personal user data which, in turn, would render a *session-based filtering approach* less than optimal.

As has been established, the recommender system will need visitors to achieve certain criteria, in order for it to categorize them into one of the three established *presentation domains*. For the domain of *Cybersecurity*, two groups can be defined with distinct criteria. The first defined criterion depends on the time that a given visitor engages with a piece of content. This criterion would make sense to add for *articles* and *videos* as these two types of content are placed statically in the domain and do not require the submission of a webform. As such, it is possible to measure the time of a given session that the user is participating in, or the length of time a visitor has played a *video*. If a visitor only engages with the content for a few seconds, then they will not achieve the criteria. If they engage with the content for a longer period of time, such as a minute, they will achieve it. The exact time a user has to engage with content needs to be defined for each specific piece of content. This is because, with the example of a video, the length of the video may be many minutes, or just a few seconds. It will be up to the content creators to define the exact time criteria for content.

The second defined criterion depends on the visitor's actions of filling out a form. Content like *webinars*, *conferences,* and *whitepapers* all require the visitor to fill out some information in a form that is placed on the webpage. This must be completed in order to download the content or sign up for the event. In most cases on nnit.com, the action of filling out a form to download a piece of content, or signing up for some type of event, takes the user away from the website. When the visitor downloads a *whitepaper*, the paper is presented as a PDF. When the visitor signs up for a *webinar*, they are taken to gotowebinar.com, which is the provider and host of webinars that NNIT uses. As such, the act of recommending content to a visitor in the same session becomes challenging. However, as the user has filled out a form, they have provided some of their detailed personal information like name, email, phone number etc. This means that the system will still be able to recommend content to a visitor based on the actions of a specific session. As an example, this could be accomplished by sending out pre-developed emails about content, related to the *presentation domain* of the registered session.

11.3.2 Structuring Data

New practices of data structuring that are brought forth by the recommender system primarily consist of the development of the three *presentation domains*. As long as the data, which the predefined presentation domain is based on, matches the types of visitors in the domain, the structuring of data that is already in place will be sufficient for this iteration of the system. If the concept is to be scaled up to a more expansive domain at some point, the practices for structuring data that are in place with this iteration would still be sufficient. The data collection channels collect data for the whole of nnit.com's areas and not just *Cybersecurity*. As such, besides the need for the structuring of more *presentation domains*, the practices that NNIT are using would remain the same.

As the current iteration of the concept utilizes a *session-based filtering* approach, a database would also be needed to structure and store the data that is provided by each session. This is important in order for the recommender system to filter visitors into the established presentation domains. Content is manually categorized into one of the three presentation domains. Users are also categorized into the presentation domains, as a result of their interaction in a session. This creates profiles that contain the same criteria as the presentation domains they have been connected to.

11.3.3 Tracking Data

Tracking of data, in this case, is important for a couple of reasons. If NNIT wants to change the purposes for their *Cybersecurity* domain at some point, or if users change their purpose for visiting the domain, then a re-evaluation of the current *presentation domains* would be needed. This re-evaluation should be based on new user behavior data. As such, if this scenario is a possibility, then NNIT will need to track users' behavior, to keep the established *presentation domains* up to date with the newest data.

Regarding *popularity filtering*, it is also important to track the number of times criteria are achieved for a given piece of content. This is important, as the system needs to be able to deem which pieces of content are more "popular" than others.

Regarding the future of the concept, the users' behavior towards the recommendations they are presented with should be examined closely. If it is found that users rarely interact with the recommendations, then the whole system must be re-evaluated. It would be a challenge to specifically define why users are not interacting with the system. It may be because the established *presentation domains* are too broad, and as such, recommendations are not personalized enough. It could also be that the way users are presented with content might not be intriguing enough. Then it could be relevant to examine or re-evaluate the users' *motivation* and *ability* regarding interaction with the recommended content. There may be many different reasons for the problem if users end up not engaging with the recommended content. However, proper tracking and timely re-evaluations of the domain's and the users' different purposes, would allow the people responsible for maintaining the system to more easily define the problem and readjust the system more effectively.

In summary, the method that the CoMa-department should use to collect, structure, and track data, regarding the current iteration of the concept, has now been established. Regarding the *collection of data*, the domain is already gathering data that allows for the three main *presentation domains* to be defined. The recommender system itself will need to gather data on a specific visitors' achievements of specific criteria in a session and then categorize the user into one of the three established presentation domains. Regarding *structuring of data*, the three *presentation domains* will have to be formed. The *presentation domains* will have information related to them, that defines the characteristics of users placed in a given group. This information will involve:

- Definition of the visitor's industry, and how it correlates to cybersecurity.
- Definition of the visitor's job role and how it relates to cybersecurity.
- Definition of the visitor's professional function.

- Definition of the visitor's authority to be a decision-maker regarding choices of cybersecurity in their organization.
- Definition of the visitor's technical knowledge in the field of cybersecurity.

Regarding tracking of data, the purposes of the domain should be maintained and evaluated, to ensure the established *presentation domains* match the original intent. Users' behavior towards presented recommendations should also be tracked to ensure a match between user and recommended content.

11.4 Step 3 - Formal Ontology

The OWL-file for the ontology can be found in Appendix 1.

11.4.1 Ontology Development

With the domain having been analyzed, a design for the recommender system having been formed, and data handling practices having been analyzed, all of the acquired knowledge can now be utilized in the development of a formal ontology.

11.4.1.1 Step 1 (Determine domain and scope of the ontology)

The first step needed in the development of the ontology is to determine the specific frames and boundaries of the ontology. This is accomplished by answering four questions, which will specify the direction of the ontology.

11.4.1.1.1 What is the domain that the ontology will cover?

The domain of the ontology is the same as the domain for the practical case in general. This means that the domain of the ontology is the area of *Cybersecurity* on nnit.com with regard to recommendations of content.

11.4.1.1.2 What are we going to use the ontology for?

The general use of the ontology will be for it to answer questions regarding which content a given user of a session should be recommended. Specifically, the people responsible for *Cybersecurity* will receive insights into which content to present for a given visitor, by being able to categorize visitors based on the content they interact with. As a result of the ontology, the CoMa-department itself will also gain insights into a new utilization of their user behavior data, as the data becomes useful in defining user groups for content recommendation.

11.4.1.1.3 What types of questions should the information in the ontology provide answers for?

The questions that the use of the ontology should provide answers for are competency questions. These questions give a general idea of the possible insights that the ontology should be able to provide. The following questions are examples of what a finished ontology for the domain will be able to answer:

- Which visitors have downloaded a *Whitepaper*?
- Which visitors have read an *article*?
- Which of the three presentation domains should the visitor be recommended content from?

11.4.1.1.4 Who will use and maintain the ontology?

The people that are responsible for the area of *Cybersecurity* will be the ones to maintain and use the ontology. These are the employees in the CoMa-department that will have specific use for the answered questions regarding the ontology's domain. However, the ontology is to be considered a proof of concept, in the sense that the final ontology would encompass all elements of the *Cybersecurity* domain. The people who are responsible for other areas of nnit.com should also have some interest in the ontology, seeing as it is designed with expansion to other areas in mind. The use of the ontology, and the maintenance of it, would be relevant for those same employees if the domain of the ontology is ever expanded to include other subdomains of the website.

11.4.1.2 Step 2 (Consider reusing existing ontologies)

In the domain of *Cybersecurity* or just nnit.com in general, no existing ontology is available to base the new ontology on. There are existing templates of ontologies as with the case of The Recommendation Ontology 0.3 (Ferris, Jacobson, 2010). However, as the ontology for this case acts as a proof of concept, it is expected to be very specific and basic in its nature. As such, a template is not optimal.

11.4.1.3 Step 3 (Enumerate Terms)

In order to start the development of the ontology in Protégé, a base of terms needs to be enumerated for use in the ontology. These terms are to be used for the different elements in the ontology. In order to strategically enumerate terms, a content audit for the terms is developed (Hedden, 2010). The content audit consists of three main rows: *Concepts*, *Variations* and *Categorization*. The terms that are defined in the first row (*Concepts*), are words that are present in the domain of the ontology. The words that are present in the second row (*Variations*) are, as the name suggests, variations for the words in the first row (*Concepts*). These variations allow for a better understanding of the original concepts and specificity in the ontology. The third row (*Categorization*) is purely for the development of the ontology itself. By categorizing the enumerated terms into specific categories, the process of naming elements in the ontology will become easier. The categorization that is shown in the *content audit* is not necessarily going to be shown in the ontology.

Table 2: Content Audit – NNIT.

Concepts	**Variations**	**Categorization**
Layperson	Top-level	User
Industry associate	Mid-level	User
Decision-maker	Bottom-level	User
Article	Readable content	Content
Webinar	Recorded event	Content
Podcast	Audio content	Content
Video	Watchable content	Content
Health check	Assessment	Content
Conference	Event	Content
Whitepaper	Downloadable content	Content
Brochure	Catalogue	Content
Brief	Introductive content	Content
Visitor	Individual	User
Criteria	Standard	Descriptor
Industry	Business	Descriptor
Job role	Occupation	Descriptor
Function	Task	Descriptor
Technical capability	Ability	Descriptor
Thought-leader	Knowledgeable expert	Descriptor
Specialist	Expert	Descriptor
Buyer	Customer	Descriptor
Data collection	Information gathering	Action
Download	Retrieve	Action
Read	Scan	Action

11.4.1.4 Step 4 (Define classes and class hierarchy)

For the structure of the class hierarchy, a top-down approach is utilized (Figure 34). The top level consists of three classes: *Content*, *PresentationDomain* and *User*. The *Content* class is representative of all content in the domain. The class has different subclasses that are representative of the types of content that are present in the domain: *Article, Brief* and *Whitepaper*. These subclasses contain subclasses of their own, which are representative of specific pieces of content: *VRCybersecurityTraining, ANewThreatLandscape* and *IAMWhitepaper*. The *PresentationDomain* class is representative of the different domains of recommendation, in which a user can be shown content by the recommender system. These presentation domains are shown as subclasses named: *BottomLevel, MidLevel* and *TopLevel*. The *User* class is representative of all the potential visitors that are to interact with *Cybersecurity* on nnit.com. Three examples of this are shown in the ontology as subclasses named: *User1, User2, User3*.

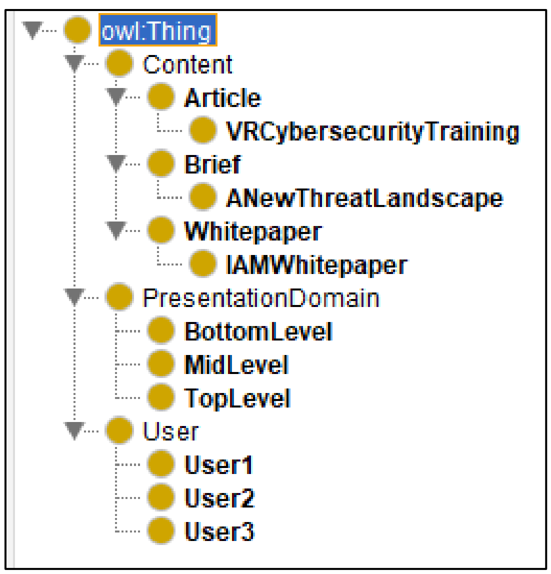

Figure 34: Taxonomy of the class hierarchy – NNIT.

11.4.1.5 Step 5 (Define the properties of classes)

The relationships between elements in this ontology (Figure 35), are solely defined with object properties. Four object properties are determined: *belongsTo*, *hasDownloaded*, *hasRead,* and *isPartOf*. The object property *belongsTo* is given higher authority, while the other three object properties are set as subproperties of *belongsTo*. The reason for this is that *belongsTo* is set as a transitive property. The object property *hasDownloaded* requires that the *User hasDownloaded Content*. This establishes that a visitor has achieved one of the criteria that allows them to receive recommendations of content from one of the three presentation domains. The object property *hasRead* requires that the *User hasRead Content*. This establishes that a visitor has achieved one of the other criteria that allows them to be recommended content from one of the three presentation domains. The object property of *isPartOf* requires that *Content isPartOf PresentationDomain*. This showcases that a piece of content is related to a given presentation domain.

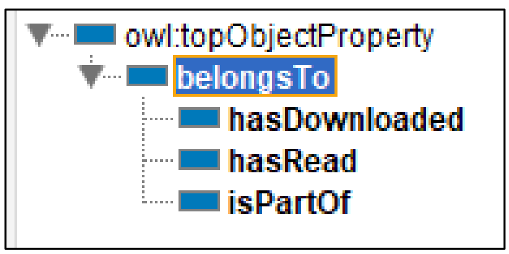

Figure 35: Hierarchical listing of object properties – NNIT.

Furthermore, annotations have been developed for the purpose of providing contextual knowledge to the object properties (Figure 36).

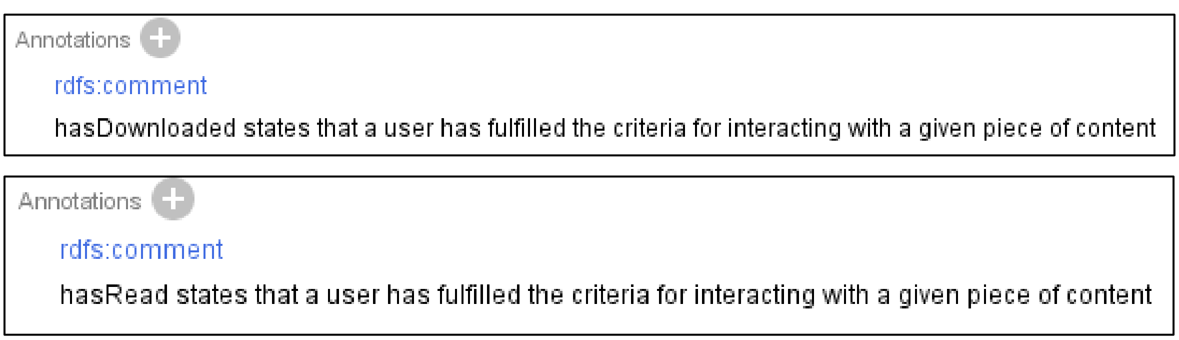

Figure 36: Annotations describing the workings of criteria through object properties - NNIT.

11.4.1.6 Step 6 & 7 (Define the facets of the properties & create individuals)

The ontology will succeed in its development without the use of facets of properties or individuals. Of course, facets of properties and individuals may indeed have a purpose in future iterations of the ontology. However, for the purpose of answering the established competency questions in this iteration, facets of properties and individuals are not needed. An example of this, in the case of facets of properties, would be the addition of cardinality restrictions. Although this would give the ontology the option of defining the priority of elements, the purpose of the ontology does not require this feature. With individuals, the required elements of the ontology are already represented through classes and subclasses. Adding unnecessary complexity to the ontology at this iteration would be detrimental to the perception of the ontology.

11.4.2 Presentation of Logic

Now that the ontology has been developed, its logic can be discussed through a visualization of the ontology itself (Figure 37). The overarching logic of the ontology is that visitors acquire the characteristics of a presentation domain by interacting with a piece of content that is already related to the given presentation domain. This is what dictates that a piece of content, which inherits these acquired characteristics, can be presented to the user, as both the user and the content are now part of the same presentation domain. This logic can also be formalized through the ontology.

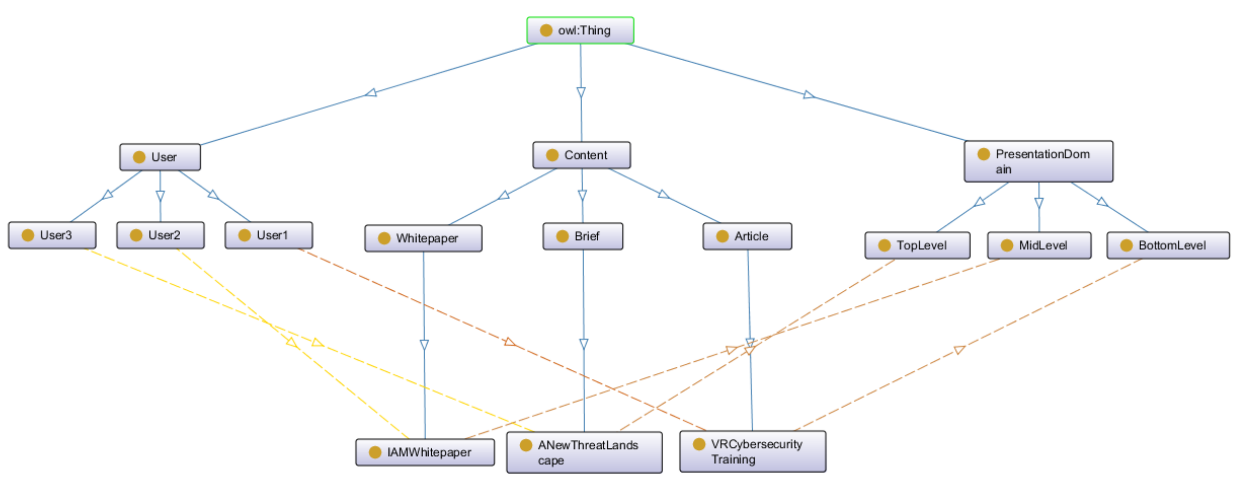

Figure 37: OntoGraf visualization of the ontology – NNIT.

The overlying logic is formalized through the statement:

A = B

B = C

Therefore: A = C

This is basically the case with transitive relationships in the ontology. The object property *belongsTo*, was set as transitive, and all other object properties are subproperties of *belongsTo*. This means that the same logic can be shown in the ontology (Figure 38)

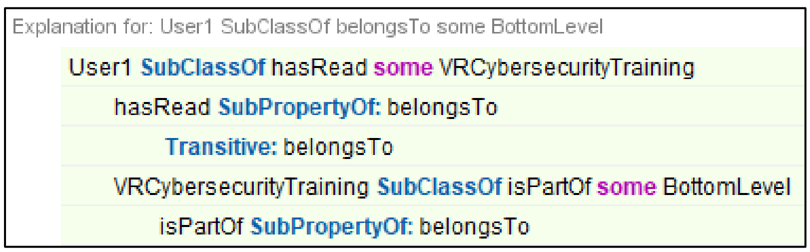

Figure 38: In-depth explanation for reasoning – NNIT.

User1 reads the article *VRCybersecurityTraining* and achieves the criteria for the content, as shown by the object property *hasRead* being connected to *User1*. The object property *hasRead* is a subproperty of the object property *belongsTo*. *belongsTo* is transitive. *VRCybersecurity* is also established as a part of the class *BottomLevel*. Since the object property *isPartOf* (that is used to connect the *VRCybersecurity* and *BottomLevel*) is also a subproperty of *belongsTo*, a coherent link is now established between *User1* and *BottomLevel*. As such, *User1* can be argued as being connected to the presentation domain *BottomLevel*. Outside of the ontology, the characteristics that *BottomLevel* contains (Figure 39) can be argued to be acquired by its connected users. As *User1* is part of *BottomLevel*, the recommender system will know the user's characteristics (because it knows the characteristics of *BottomLevel*) and therefore be able to present the user with similar content that is part of the presentation domain *BottomLevel*.

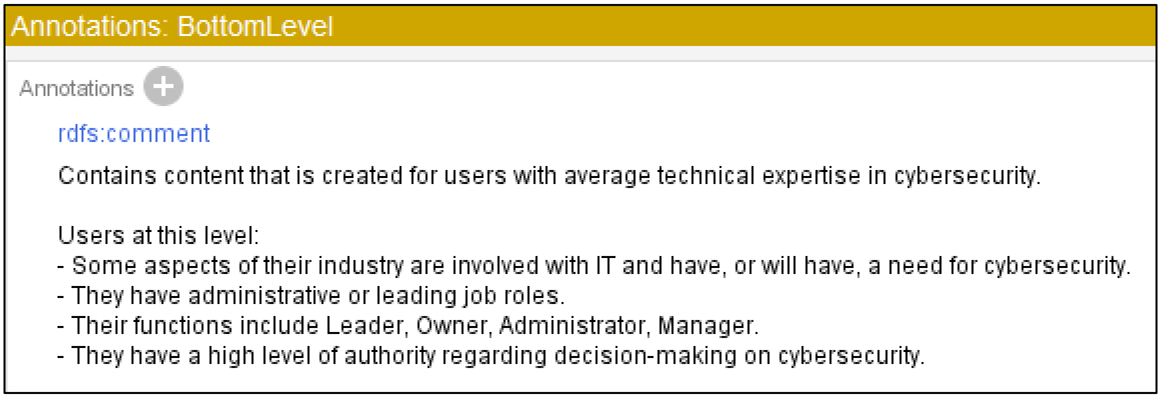

Figure 39: Annotations describing the characteristics of the BottomLevel class – NNIT.

11.4.3 Answering of the Competency Questions

With the transitive logic of the ontology explained, the ontology should be able to provide answers to its defined competency questions. This ultimately showcases the value of the ontology.

Competency question 1: *Has a visitor downloaded a whitepaper?*
The ontology states that: *User2 hasDownloaded some IAMWhitepaper* (Figure 40). As such, the criteria for engaging with a downloadable piece of content have been met, and it must be true that a visitor has downloaded a whitepaper.

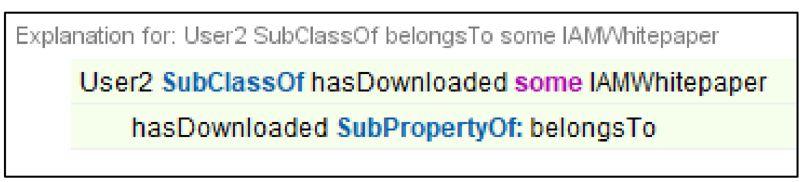

Figure 40: Formalized explanation related to competency question 1 – NNIT.

Competency question 2: *Has a visitor read an article?*
The ontology states that: *User1 hasRead some VRCybersecurityTraining* (Figure 41). As such, the criteria for engaging with a readable piece of content have been met, and it must be true that a visitor has read an article.

```
Explanation for: User1 SubClassOf belongsTo some VRCybersecurityTraining
    User1 SubClassOf hasRead some VRCybersecurityTraining
        hasRead SubPropertyOf: belongsTo
```

Figure 41: Formalized explanation related to competency question 2 – NNIT.

Competency question 3: *Which of the three presentation domains should the visitor be recommended content from?*

The ontology states that: *User2 hasDownloaded IAMWhitepaper* (Figure 42). *IAMWhitepaper isPartOf MidLevel*. As such, by interacting with a piece of content from the mid-level presentation domain, and achieving the respective criteria, it is shown that it is the mid-level presentation domain that applies to the visitor. Therefore, the visitor should be recommended content from the mid-level presentation domain.

```
Explanation for: User2 SubClassOf belongsTo some MidLevel
    User2 SubClassOf hasDownloaded some IAMWhitepaper
        hasDownloaded SubPropertyOf: belongsTo
            Transitive: belongsTo
        IAMWhitepaper SubClassOf isPartOf some MidLevel
            isPartOf SubPropertyOf: belongsTo
```

Figure 42: Formalized explanation related to competency question 3 – NNIT.

11.4.4 Interaction Graph

In order to visualize how the ontology and the workings of the recommender system are intertwined, an *interaction graph* has been developed (Figure 43). The visualization showcases how a user interacts with a specific piece of content (by fulfilling the given criteria). This piece of content contains a content profile which tracks the number of times its criteria have been achieved. The piece of content is also part of a presentation domain which contains characteristics. These characteristics are given to the user, who interacted with the specific piece of content, and are added to their user profile. The presentation domain also contains some recommendation options (content), which are similar to the specific piece of content that was originally interacted with. These recommendation options have been interacted with by other users, who themselves have received characteristics from the presentation domain. The original user is then recommended content from the recommendation options.

Going back to the discussed point of expanding the recommender system to other areas of nnit.com, users will have the option of being part of more than one presentation domain at the same time. A way to utilize the ontology in this case, would be to establish logically defined classes. These defined classes would basically be the user profile (defined by the dashed circle in the *interaction graph*), containing the characteristics of all related presentation domains.

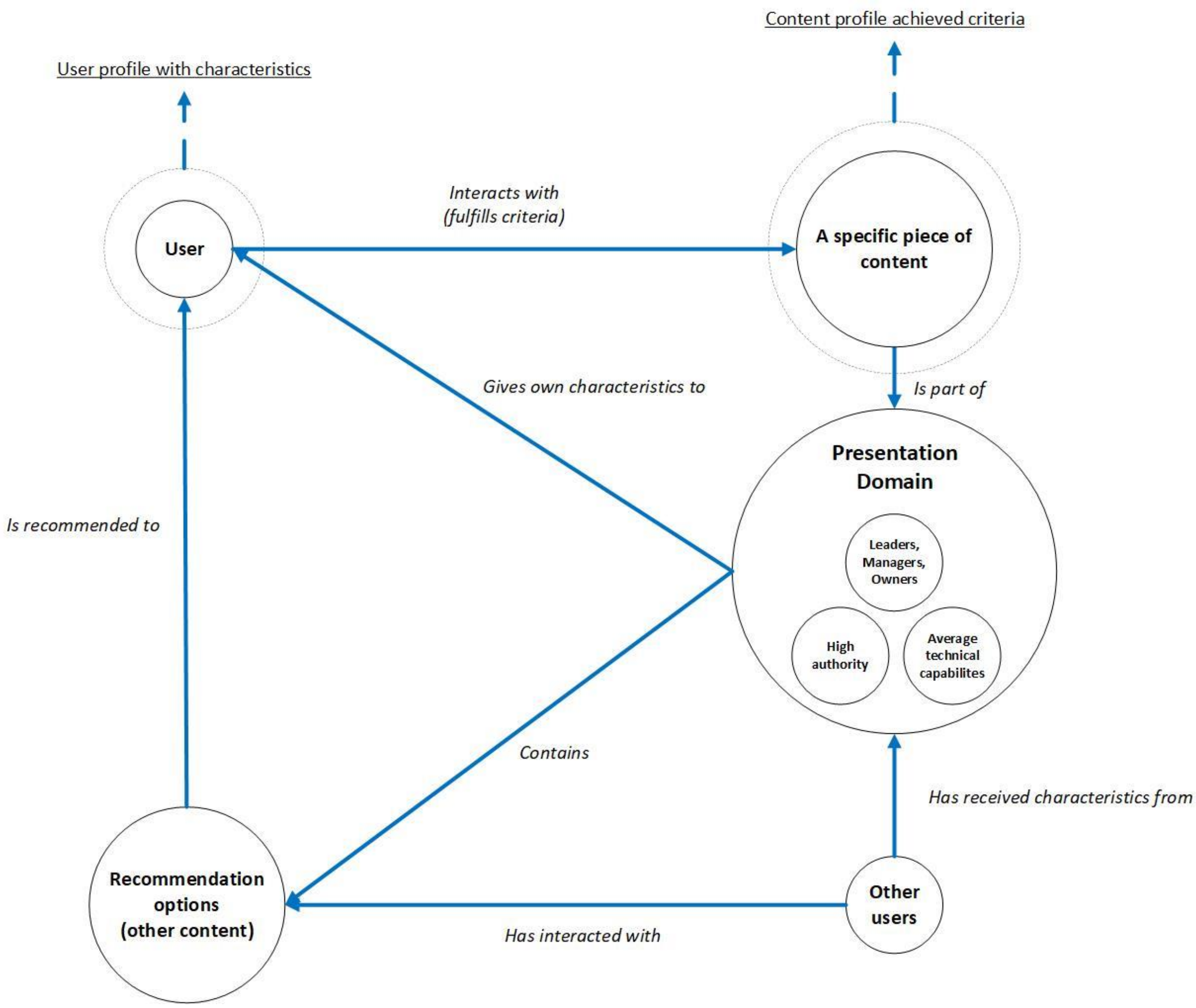

Figure 43: Interaction graph of recommender system – NNIT.

11.4.5 Ontology Evaluation

With a formal ontology that answers the question of which presentation domain a given user should be presented to, the recommender system will now be able to easily sort and categorize users visiting *Cybersecurity* on nnit.com. This establishes the transitive logic of the proof of concept. In its full development, every piece of content existing in the area of *Cybersecurity* needs to be added. More data that is related to the content may also be added. An example of these types of data would be different languages, author, IP etc. Although this is not needed for the current iteration of the ontology, it may be relevant to add more knowledge through data in later iterations. The specific users themselves will stay as placeholders until the ontology is implemented in practice and able to utilize real visitor-sessions. Now that the early proof of concept has been established, the ontology can be improved upon by not only defining a relatively large group of content from which to give recommendations, but also delve into the visitor's motivation for interacting with a piece of content.

11.5 Step 4 - Persuasion

The current iteration of the system recommends content based on *session-based*, *collaborative*, and *popularity filtering* approaches. With this *hybrid filtering* approach, the filtering is already substantial. However, the system can become better by introducing persuasion through B.J. Fogg's Behavior Model. The goal with utilizing the Behavior Model will be to ensure that the system is risk-aware for its users.

11.5.1 Behavior Model

The first aspect to be examined is *ability*. In this case *ability* refers to the visitor's technical ability to understand and use a specific piece of content. *Ability* is already accounted for, as visitors are already categorized in one of the three presentation domains, with each presentation domain having a defined level of technical ability needed, in order to gain value from the content that is presented. As such, visitors will have no trouble regarding their *ability* to understand and make use of the content they are presented with.

Motivation is the next part to examine. *Motivation* in this case refers to how motivated a user is to interact with the content that is recommended. This is important, as the many different types of content that are present in *Cybersecurity* demand different levels of motivation. A *video* or an *article* on nnit.com may only take a few minutes to interact with, and as such, do not require much *motivation*. However, participating in a *webinar* or a *conference* are examples of content that demands a lot of *motivation* from the visitor, as the content takes a lot of time and effort to interact with (participate in). Currently, there is not an option in the system or ontology to decipher the visitor's *motivation*. However, this could be created with a similar solution as the tracking of achieved criteria for a specific piece of content. As the recommender system is *session-based*, the tracking of the visitor's *motivation* must come from their interaction with the content in the specific session. As such, all content in the domain will receive a level related to the amount of *motivation* that is expected to be needed in order to interact with the content.

Level 1 will be given to content that only needs a low level of motivation, like *articles* or *videos*. *Level 2* will be given to content that needs a more substantial level of *motivation*, like *briefs* or *whitepapers*. *Level 3* will be given to content that needs the highest level of *motivation*, like *webinars* and *conferences*. This score will have to be given manually to each piece of content by the content creators, as content of the same type may still differ drastically in the amount of *motivation* needed, like 30-second *videos* or 5-minute *videos*. The *motivation* level for content will be stored in the content's profile, just like the number of times its criteria have been achieved. The visitor will also have to gain a *motivation level* for their user profile. This level remains equal to the piece of content the visitor has interacted with in the session. If a visitor achieves the criteria for reading a short *article*, their *motivation level* is set to 1. If they then download a *whitepaper* in the same session their *motivation level* will increase to 2. While this happens, the recommender system will adjust itself to recommend content that fits with the *motivation level* of the user. In the ontology, the *motivation level* will be represented with annotations for the content and users (Figure 44).

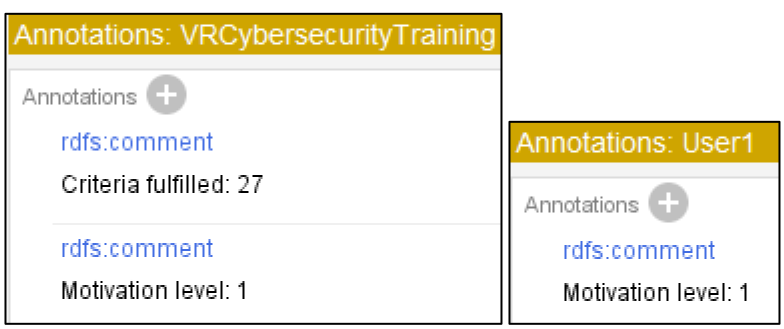

Figure 44: Motivation levels represented with annotations in the ontology – NNIT.

As has been mentioned, a lot of the content in the domain is either downloadable or takes the visitor to another page to accept an invitation. One would assume that since the recommender system is *session-based*, the session ends when a visitor fills out the webform and downloads the content or is taken to a new website by the link. This is also why, with this type of content, further recommendation needs to occur over email after the visitor has provided it. However, when a user downloads content or is taken to another site to accept an invitation or similar, the session on nnit.com has not necessarily ended. Often, a new browser tab is opened when being transferred to a new site. As such, the current session is still intact. Even though it is seldom the case, a visitor still has the opportunity to continue their session in the nnit.com tab. Therefore, the introduction of *motivation levels* still impacts these visitors, even though it will mostly impact visitors who engage with content that has its main presence on nnit.com, like *articles* or *videos*. The *motivation level* will of course have to be weighed against the already existing *popularity filtering* that is occurring simultaneously. Overall, the system will end up recommending the visitor with content that has been interacted with by similar users (the same presentation domain), but also content that has been defined as the most popular content of the recommendation domain, that also fits the visitor's *motivation level*.

Prompts are the last part to examine. The relevance of the recommendations is already ensured through the different filtering-approaches that occur. As such, the purpose of *prompts*, in this case, is to show transparency to its visitors. This is accomplished by modifying the visual recommendations on nnit.com (Figure 45). Through the use of *spark prompts,* descriptive texts that explain the expected amount of time a piece of content takes to finish should be presented. In the *recommendation window*, *videos* should showcase their length and *articles* should describe the expected amount of time it takes to read it. This way of describing the amount of resources (time) a visitor has to put into a specific piece of content, assists in establishing a more transparent relationship between the visitor and the domain. This might incentivize the visitor to engage with a piece of content, knowing the time of interaction.

Figure 45: Example of final recommendation – NNIT.

With *ability*, *motivation,* and *prompts* having been accounted for, the system is set to persuade the visitors of the domain and increase the probability of possible interactions with recommendations. For the CoMa-department most of the aspects of risk awareness that can be adjusted are done so with the Behavior Model in mind. The system ensures that visitors have the ability to interact with the content they are recommended, they are expected to have the level of motivation needed, and they are prompted to engage through transparency. Remaining aspects of risk awareness to be considered are the rhetoric, design, and stability of the recommendations. As the recommendations are to be placed on an already existing platform, it is a given that the mentioned aspects should conform to the platform they are a part of. As such, the recommendations will be non-intrusive and keep the same rhetoric, design, and stability that visitors are already familiar with from the rest of the website.

11.6 Evaluation

With all four major steps of the concept having been developed for the case, the design for the recommender system is concluded.

The system sets out to provide relevant recommendations for all the different kinds of users that visit *Cybersecurity* on nnit.com. By delving into the domain's context, content, and users, a match was found between the purposes of the domain, the content that is present on the domain, and the visitors' purposes for visiting. The recommender system is set to utilize a *hybrid filtering* approach. *Session-based filtering* and *collaborative filtering* are utilized in order to filter visitors and focus on their purposes for visiting.

Popularity-filtering is introduced to further specify and present more relevant content to the visitors. *Collection*, *structuring*, and *tracking* of data were mapped in order to define the utilization of data that would create an adequate foundation of data for the recommender system. Herein lies the criteria that visitors need to achieve in order for the system to recognize an interaction. Then the semantic structuring of the information was accomplished through the development of a formal ontology. This proved able to provide answers to relevant questions in relation to recommendations, showcasing the ontology's explanations and the logic of the system at the same time. Finally, the persuasion of the system was designed through the utilization of B.J. Fogg's Behavior Model, and the system's risk awareness was also examined. This ensures that the recommendations themselves fit the visitors' *ability* and *motivation*, prompting them to engage with the recommendations. As a final touch, the recommendations rhetoric, design and stability were examined and defined to conform to nnit.com's identity.

The people responsible for the *Cybersecurity* domain in the CoMa-department have voiced their interest in the design of the system and agree that it has practical value for nnit.com and their department. For them to introduce such a system to their platform, it would need to be expanded to encompass more of nnit.com, or at least the complete area of *Cybersecurity*. Fortunately, there is no questioning the system's potential to be expanded to the desired size and responsibility. By developing the system through these steps of the thesis' theoretical concept, and building a formal ontology explaining the logic, the people responsible for the domain have gained new insights into their data, visitors, and communicative purposes of the domain. As such they are satisfied with the result of the practical case.

11.7 Comparison

Now that the two practical cases have been concluded, a comparison between the two is possible. By comparing the two cases' similarities and differences, the value that is created through the use of the concept will be examined. This will allow for research-based interpretations and conclusions to arise, providing answers to the original problem and hypothesis that the thesis set out to investigate.

11.7.1 OpdagDanmark

OpdagDanmark's recommender system reflects the position of a startup company that has not properly established their data potential. OpdagDanmark is not a big competitor in the tourism market yet, which means that they are not a go-to solution compared to more established brands like TripAdvisor. They still need a proper foothold in regard to monetization and more time to properly establish their product. The data that has been used for the recommender system has been sparse, and the final concept idea undeniably reflects this aspect. However, the proof of concept also indicates the potential that lies in expanding upon the initial concept idea, which ensures value throughout the future work of developing their app and growing their user base. It is only a matter of time before the relationship between the application and the concept can move into a strong, intertwined position of practicality, using the academic principles from the steps to form the practical utilization of the data. In fact, the recommendation aspect can help enhance their product extensively.

Currently, OpdagDanmark offers an active explorative experience in their app to find and learn more about experiences in Denmark. The recommender system gives this aspect a new layer by adding a service that can point users towards interesting experiences without having to actively search for them. This essentially builds upon the user experience and the value aspect that the users are getting from the app, and it can even be an important factor in getting new users into the domain. This boosts interaction and can have an impact on conversions for premium content, if users are getting value from the improved user experience.

If the users know about the capabilities of the app, and actively choose to seek recommendations from the recommender system, they are also willing to share data in order to receive the best recommendations. This can help develop the user profiles and make more personalized recommendations. Over time, this part of OpdagDanmark's product can become its own selling point, becoming a must-have companion for spontaneous travelers all around the country. This may create some interesting job opportunities for Information Architects in regard to maintaining and developing this system with the available data. There is always room to expand upon the knowledge aspect of users, gaining a deeper understanding of their habits and preferences. Therefore, this might be considered a catalyst for creating information-based jobs in the company and add to the overall practicality of information architecture as a worthwhile investment for data-driven businesses. For OpdagDanmark, this can position them uniquely on the market, where the product that they offer can further stand out from the current options.

The personalized recommendation might even become the main product at some point, further servicing users who have learned how the system gains more knowledge about them. This could be through utilities such as experience logging. This essentially lets the users utilize the recommender system actively, as they input relevant data that they want the system to consider in a recommendation. This could be implemented as a sort of filtering/search option in the overall application, which would bypass the risk awareness of the recommender system. It might be able to be implemented using the same established ontology and data handling practices that have been examined throughout the steps. This will essentially show how the concept can be the foundation for brand new features in OpdagDanmark's overall domain and product, as long as it revolves around the relationship between users and content.

11.7.2 NNIT

The recommender system that has been designed in the case of NNIT (specifically the CoMa-department) brings forth different practical usages, and as such, practical value. On a basic level, the introduction of a recommender system, presenting recommendations through a new element on the website, provides an added aspect to the existing user experience. The system adds to a more exploratory user experience through the recommendations themselves. The recommender system also provides new insights for the people in the CoMa-department who are responsible for nnit.com. By introducing automated recommendations to visitors of the defined domain, the employees who maintain the domain are forced to evaluate their strategic communication. The system is focused on creating three presentation domains that connect the domain's purposes, the content that is present in the domain, and the users' purposes for visiting. By developing a system that attempts to match recommendations of content between all these aspects, the CoMa-department gains insights into their external communication. From these insights, they are better suited to answer whether their goals for the domain are appropriate and are being fulfilled with the current communication. Finally, through the development of a strong semantic information architecture, the CoMa-department is also given insights into how they can utilize their data, with respect to visitors and content, for the recommendation services.

The recommender system, as it is designed for this case, is a proof of concept. While the system is designed to be able to encompass all of the *Cybersecurity* domain, it has the potential to expand to the rest of nnit.com. Modifications could also be made to the system to make the recommendations more personalized. If more features that collect data are added to the website and the system does not have to rely on specific sessions for recommendations, more personalized recommendations may emerge. However, the current design of the system is a result of its current information environment and data handling practices. For some of the content, the goal is definitely to persuade the right individuals to choose NNIT as a provider. However, the specific content that is recommended is not a product for customers to buy. It is meant to build trust, inspire, and showcase NNIT's expertise. As such, modifications and data collecting features are advised to not be developed solely for the purpose of the recommender system. With the current designed system, the recommendations that are provided are not personalized to each specific visitor (besides with the introduction of the *motivation level*) but rather focused on one of three large presentation domains (user groups). However, the system manages to do this while keeping in line with the existing data collection and identity of nnit.com.

The value of the current recommender system, and its potential for expansion, sets the agenda at the CoMa-department for a future with a stronger focus on nnit'com's information architecture. As a result of a better understanding of the capabilities of a system like the one in this case, NNIT may have found more reasons to utilize the semantic potential of their data through a semantic layer, which provides logical insights for maintainers, and which may assist in ensuring a transparent human-to-human interaction between the users and the company.

11.7.3 Comparison between OpdagDanmark and NNIT

OpdagDanmark is a start-up and NNIT is an established organization. The differences between the two cases have been clear from the start. Working with OpdagDanmark's data handling in order to build a strong recommender system has proven to be a case that has the potential to evaluate and create a foundation for the company's core principles regarding its utilization of user data. With NNIT, the recommender system only ever reaches the platform of nnit.com (even when expanded), as the case is focused on a website domain, rather than the whole of NNIT. As has been found, this difference in purpose and potential of the two systems results in vastly different types of values that can be gained from each case. As such, the two cases are not comparable when discussing if one system is more successful than the other. The two designed systems accomplish their goals and, as such, fulfill the purposes that were defined by the initial meetings. As such, it is clear that the two recommender systems do create an aspect of practical value for each of their respective cases. If these two vastly different cases can provide practical value for their respective companies, from a recommender system that is built through a similar approach to information architecture, it has the potential to create practical value in data-driven companies. As such, an argument is made for the concept's flexibility, as it is able to be replicated in many different scenarios. This further points towards the knowledge, principles, and tools of a semantic information architecture being valuable and useful in creating practical value (logical insights) through its distinct dissemination of data.

11.7.4 What can we take away from these cases?

As the problem that the thesis attempts to address has consisted of businesses not utilizing the full potential of their information environment, and the hypothesis stating that principles of information architecture can assist in improving companies' data utilization, the question now is:

Does the thesis scientifically showcase the practical value of an information environment that utilizes its semantic data potential as a result of the development of a semantic layer?

Combining the findings of the two cases, the short answer would be "yes". However, an explanation is needed in order to fully grasp how this is accomplished. The developed concept sets out to design the optimal recommender system for its given information environment. It also has a core goal of showcasing how this can be accomplished by utilizing a semantic layer, exhibiting the practical value of utilizing data semantically. This ensures that a competent recommender system is designed, while also providing awareness of the potential and need for information architectural change in future expansions of the system. The concept provides a repeatable framework while utilizing principles of information architecture.

With the focus of ultimately developing a recommender system, an analysis of the domain is made in *Step 1*. With the Information Ecology Model, the business of the case is prompted to evaluate their *context*, *content* and *users* in this domain. In *Step 2*, they must evaluate their data handling principles of the domain. By investigating their current use of syntax data, the value that is gained through the insights may even reach beyond the domain of the recommender system. The case with NNIT is a clear example of this. With the concept prompting the employees in the CoMa-department to evaluate the communicative purposes of their platform (nnit.com). With *Step 3* and *Step 4* of the concept, the logic of the system is proven, its risk awareness is optimized, and its communication is accounted for. These steps do not only showcase the theoretical success of the system, but they also provide clients with a clearer understanding of their information environment related to the recommender system. The sequence of steps also proved valuable in structuring the process of the concept, since they were placed in a beneficial way for the concept's flexible nature.

An interesting take-away from working with a recommender system, is the nature of the type of problems that a recommender system attempts to solve. A recommender system will always deal with the anticipation of user's (people's) actions. This implies that the problems a recommender system attempts to solve can be considered *wicked problems*. While *wicked problems* are difficult to define, it can be stated that they have no right or wrong answer, just successful and unsuccessful ones. The effectiveness of the solutions to *wicked problems* is also difficult to measure, since it is based on the human psyche, which is ever-changing (Strategy as a Wicked Problem, 2014). While these are just a few examples of what defines a *wicked problem*, it is clear that recommender systems deal with the same challenges as listed. The concept's attempt to build the most optimal recommender system is achieved through the use of structuring data semantically. However, this is still utilized to anticipate user's (people's) interests and interactions. As such, interpretations will always be needed when defining the future success of data that is transformed into qualified guesses (the concept of: "*to recommend content*"). This is why its success can only ever be theoretical until the system is implemented and tested in practice. To highlight the concept's practical value regardless of testing and implementation, as explained with *Step 1* and *Step 2*, the insights that clients receive through domain analysis and data handling is knowledge that can be utilized for other practical purposes.

In order to define the success of the thesis in providing answers for its problem and hypothesis through the established concepts, the companies have to start to utilize the semantic potential of their respective information environments. Through the thesis, it is implied that people (and therefore companies) who are involved in the cases' work processes, are working with the semantic potential of their data. This is, of course, a result of the concept forming a semantic layer through the utilization of different principles of information architecture. This accomplished in order to develop a scalable design for an optimal take on a recommender system for each company. As such, working with semantically structured data to develop recommender systems is one way in which practical value is showcased in the thesis. There is potential to implement more algorithms on top of the existing semantic layer, not just recommender systems, to showcase further value from the established semantic layer. This is also in line with the purpose of the thesis, since a recommender system is not meant as the sole product to convey. Rather, it is that companies can utilize their data semantically through a semantic layer for technologies such as recommender systems, showcasing a practical value aspect.

11.8 Discussion

This section will dive into some of the findings, implications and problem areas associated with the finished cases. The subjects that are going to be touched upon here will be regarding the role of the information architect in a practical setting, the ethics of the featured persuasion, the privacy issues associated with collected user data and how the featured concept can be (and to an extent already is) affected by the rise of AI in technology; what consequences does this bring in the future? These discussions will tie into a prediction on the future for the featured cases, and how to ensure that they will not compromise their users unethically.

11.8.1 The Role of the Information Architect

First of all, it is important to stress that the process to finalize the concept has given a lot of useful insights, especially in terms of the role of the information architect. This role is rooted in design, since there are a lot of different design aspects that play a role in an information environment. Data can be considered a sort of building block, and it is the job of the information architect to build something insightful and valuable from the applied data sources. Even though information architecture is powerful when being applied to the right problem areas, in practice, it still heavily relies on other roles to then finalize certain conceptual ideas that the design has brought forward. In order to actualize concepts, acquire valuable insights and strengthen user experience etc., it must be accomplished with a team of people that accounts for the "before" and "after" associated with a given information architecture design that has been developed.

This means that when arguing for the value of information architecture, it is important to point out supporting roles such as programmers, system managers or concept developers that are able to fully realize the design of the architecture. When arguing for the value of information architecture in a situation that can potentially lead to an employee position in a company, this is an important aspect to put emphasis on. Information architects find their presentable value through a niche and technical understanding of given information environments. As such, they can provide the company with advanced insights into the development and optimization of their business. This gives an understanding of exactly how to fit this academic field into their overall business structure and hire capable profiles on the best possible premises. This also relates to the concept featured in this thesis, as a semantic structure is part of a larger business ecosystem.

11.8.2 Persuasion

> *"The recommended product is what the user would have chosen with or without the recommendation. The system is picking the best. But, would the user really have chosen the same product without the recommendation? Do recommendations change user behavior or do they merely shorten the path of an inevitable choice?" - Katakam, 2019.*

Generally, this moral grey area that Katakam puts forward is something that a company on one hand wants to master the function behind, as it shows that they have a grasp on their users' psychology, but on the other hand it also puts forth a negative connotation tied to manipulation and exploitation. Nobody wants to be subject to accusations of manipulating their users which, as a modern example of unethical practices related to recommendations, is a growing concern in the video game industry. Here, governments, on behalf of the consumers' own request, have been forced to jump in and react to certain persuasive marketing tools, in an attempt to quell unethical practices. ("Gaming loot boxes: What happened when Belgium banned them?", 2019). The video game companies were crossing certain ethical lines between manipulation and monetization; essentially playing with gambling principles.

It is safe to say that persuasion in digital products has become incredibly efficient, but also dangerous. This is important to keep in mind with the eyes of an information architect, who is able to establish logic that argues for persuasive content. If there is no ethical filter to future persuasive technology, the digital climate can become dramatically dangerous for the mind.

Thankfully, there is no money involved as of right now in both cases featured in this thesis, and it is open to discussion if the precautions caused by altering the choice of both companies' users are even worthwhile to debate from a moral standpoint. However, it is possible that in the future, the concepts' information environments may be utilized for unethical practices. The information environments are, after all, based on real people and might at some point become very detailed with information. This puts forth the possibility of gathered insights being misused and applied to unethical and persuasive means. Means that essentially prey on inferred assumptions about the users.

For NNIT, this is a very limited phenomenon, since recommending the different content that is present on their website is hard to push on users in a way that could be deemed unethical. However, in terms of OpdagDanmark's concept, this could become a real problem at one point. For example, (staying with the defined domain of pizza places) with implications of fast food addiction. Here, an example of how the unethical approach would formalize itself, could be to present the user with an excessive amount of coupons and offers, or start utilizing gamification principles (do this or that to level up and/or receive offers) regarding eating at different fast food restaurants throughout the experience map. This could foster some bad habits for certain types of people.

11.8.3 Privacy

The discussion regarding persuasion opens up into the debate concerning privacy. Privacy, especially online, has throughout the 2010's been drastically compromised by American tech giants such as Facebook and Google (Owens, 2019). Both companies sell relevant user information to advertisers that are then able to utilize the Facebook feed and/or Google search results to advertise their product to the most fitting demographic. This has proven to be an incredibly efficient monetization strategy simply brought along by harvested information. However, throughout the 2010's, big tech has been notorious for being secretive with exactly what kind of information they are harvesting. Trust in Facebook went down dramatically after the Cambridge Analytica scandal ("The Cambridge Analytica Story, Explained", n.d.) went viral and showed how big data analytics (Hilbert, 2016) could be used in the wrong hands (Weisbaum, 2018). Although not influenced by it, the EU passed a major political intervention in terms of data transparency shortly afterwards. The GDPR, as it is called, has provided a global standard in terms of online privacy, which businesses need to adhere to on a global scale by being transparent about how they collect data (Chee, 2019).

Even though trust in big tech companies has been waning after the GDPR put forth a status quo “rulebook” regarding online profiling, the COVID-19 pandemic was positive for big tech as a whole (Newton, 2020). The pandemic showed another side of companies, as suppliers of comfort and an important connectedness while people were enduring lockdowns around the world. Social media services, streaming, online shopping, information searching; all of it played a role in decreasing the sense of loneliness and despair that was imminent during a pandemic. The numbers are reflecting this trend evidently with big tech stocks skyrocketing throughout the pandemic ("How Big Tech Got Even Bigger", 2021). Big data has therefore shown its capabilities as a double-edged sword, and not just a means of advertising possibilities or political scandals. It has been at the core of how big tech has been working with governmental forces in order to get the pandemic under control (Foer, 2020). The question of whether big data is going to continue to endure criticism as a privacy-breaching practice, as it was before the pandemic happened, cannot be answered currently. The public eye might become more accustomed to the idea of big data, adopting a more accepting outlook based on its helpful role throughout the pandemic.

This is where OpdagDanmark and NNIT comes into the picture. Even though there has been great controversy regarding personal data in the late 2010’s, it seems that it might only be a question of time before the general public adopts an accepting outlook on companies using personal data in the post-pandemic world. It is no secret that, out of the two cases of the thesis, OpdagDanmark, in particular, is compromising their trust regarding the use of data to build carefully and well-constructed user profiles. However, if there is no possibility for the data to be linked to real people, and therefore completely anonymized when accessed by potential hackers, the ethical implications brought forth by a user’s preference and location sharing should be a solid case morally (Hassan, Habegger, Brunie, Bennani & Damiani, 2013). NNIT should not face problems either. With the current iteration of the recommender system, they are not basing recommendations on specific user’s personal data. Instead, predetermined groups where the users' actions are what determines their categorization is the basis for recommendations. As long as NNIT are transparent with how they are tracking their website visitors, data privacy concerns will not become a problem related to the recommender system.

Even if both OpdagDanmark and NNIT acted reticently about how they are treating collected data, the product that the companies offer might make people justify this aspect, especially if there is sufficient motivation to utilize it. This phenomenon is called the privacy paradox (Barth & de Jong, 2017). With that said, companies must still tread carefully when dealing with data, even when it is in a completely ethical way. The connotations associated with the late 2010's privacy debate (Owens, 2019) still makes it an incredibly delicate subject to communicate to users.

11.8.4 Artificial Intelligence

Finally, it is important to take a look at AI. According to Bostrom, it is inevitable that AI will overtake current technological standards since all clues point towards sophisticated AI as being a feasible project (Bostrom, 2016). In fact, this is possible in the very near future based on the idea of the singularity, which states that technological advancement progresses exponentially (Kurzweil, 2005). The purpose of the core concept of this thesis is to put forth a design for a semantic layer that both supports and accounts for an overlaying recommender system. This in itself is not necessarily AI, since there is no argument associated with automation. However, modern automated recommender systems are considered weak AI (Bostrom, 2016) which means that it adopts the idea of creating brand new information with the use of structured data. This is what both OpdagDanmark's and NNIT's concepts are capable of, and possibly how they are going to be implemented in a practical way. Having both recommender systems automated by an AI-algorithm is highly preferable and might even be considered important in order to create sufficient value alongside the potential for general insights. How this would work in practice is by having an AI-algorithm automate the recommendations themselves, but with the fundamental reasoning from the ontology and things like persuasion implemented as well in conjunction with risk awareness for example. This would create an intelligent system that could account for new knowledge and act on optimal data connections, with a semantic layer to ensure a human-like approach to logical reasoning behind the scenes. This means that the design can be considered easier for humans to decipher and understand.

One of the reasons that a semantic layer is so important in recommender systems is that it can be considered explainable AI when accounting for the possible automation (Bianchi, Rossiello, Costabello, Palmonari & Minervini, 2020). Explainable AI accounts for all relationships between every single data snippet and is therefore relevant because of the option to account for logical reasoning. Furthermore, it makes the overall domain possible to understand fully for AI-developers. This means that everything that comes out of the recommender system (if it becomes automated with AI) is completely transparent and therefore even ensures that new problematic areas such as unprecedented bias does not occur.

This is one of the obstacles with unexplainable machine learning algorithms currently, since they are reinforcing pre-existing inequality in ways that their developers do not fully understand (O'Neil, 2016). The actual work that the algorithm does in terms of data analysis proceeds in ways that developers cannot always account for, which means there is not always a logical explanation for certain unethical outcomes. An example of how these tie into recommendations comes from Facebook's group function, where users are reporting recommendations to join white supremacist and anti-Semitic groups, seemingly out of nowhere (McEvoy, 2020). This has nothing to do with Facebook's own views but is happening because the algorithm that controls the recommendations sees the groups as successful. The problem is that the logical criteria for "*being successful*" is likely not well-defined in this domain, and essentially comes down to sheer engagement. In other words, the unethical nature of the groups is not accounted for at all, because the AI cannot reason for this aspect. This ties in with the need for AI-functionality that has reasoning capabilities for why it is recommending certain content. This gives the developer the potential to dive in and block it from recommending things that can be detected as problematic. In contrast, explainable AI can therefore be considered the ethical alternative to AI-development. If anything unethical happens, which affects the communicative aspect of the user experience, it should be possible to account for it immediately. The programming therefore has to meet certain criteria of transparency (Bostrom, 2011). Once again, for NNIT, there does not seem to be any cause for concern regarding problems of this nature in their specified information environment. However, for OpdagDanmark this reasoning aspect could be relevant alongside the persuasion problem of recommending fast food en masse to inferred addicts. This should be accounted for in the algorithm as something unethical.

Semantic architecture is also starting to take hold in other AI-automated practices, as with the case of financial fraud analysis, as an example. Here, problem solving ontologies use reasoning to detect money laundering schemes and other suspicious behavior (Chmielewski & Stąpor, 2018). The transparency of logical reasoning is at the center of this architecture-design principle, which makes it versatile and highly useful on a lot of information domains, not just for recommender systems. It is therefore wise to establish an ontology as a foundation for multiple AI-designs that a business might implement to oversee multiple technological functions simultaneously on a given information domain. Utilizing semantic information architecture can therefore be considered futureproofing based on Bostrom's predictions (Bostrom, 2016).

This concludes the discussion chapter for some of the most controversial discussion points brought forth by the work with the concept and cases. To finally conclude upon all of this in retrospect, with the insights from the *Comparison* chapter, the conclusion will carefully try to amalgamate the most prominent academic takeaways.

12 Conclusion

Now that both cases have been developed and compared thoroughly with a supporting discussion of future prospects, their overall connection to the initial problem area and the hypothesis can be surmised. The problem area was specified as businesses not utilizing the semantic potential of their data. This was followed up with a hypothesis, which stated that a recommender system, designed with what turned out to be the featured conceptual steps, can produce valuable insights for a targeted information environment. This would, in turn, underline one way in which information architecture could become valuable in professional environments.

After an initial knowledge base has been established between consultants and clients, the concept attempts to collect and track relevant data, construct a formal ontology and implement persuasion principles. This is accomplished by introducing what can be considered an overarching semantic layer into the initial information environment. The concept has shown to be replicable, pointing towards the concept's flexibility and ability to conform to different information environments.

The value of the two cases is not focused on a direct monetization aspect, but are instead indirectly improving alternative means of conversion, by improving aspects of the user experience. A deeper and more logically mapped information environment is the catalyst for being able to build these improved aspects of the user experience, where a recommender system functions as a practical example of this. Although this thesis has only been able to examine two cases, with the arguments that has been presented throughout the thesis, it is deemed a valid theory that many data-driven businesses, especially those that are organizationally considered similar to OpdagDanmark and NNIT, may also be able to gain an aspect of value from structuring their data semantically. As this is not an aspect that can be directly concluded upon in this thesis, it is up to the concept's proficiency in a professional market to thoroughly test this hypothesis and investigate the concept's average aspect of practical success. This gives the concept a purpose outside of this thesis.

Precautions have been taken in terms of ethical implications that touch upon persuasion, privacy and AI, which examines the future prospects of semantic information architecture. Recommendations are risky by nature, as they can negatively impact the user experience by being intrusive. To account for the logical aspect, which ensures in-depth argumentation for specific content recommendations, the semantic layer is utilized. This is designed for, and by, humans, to ensure a domain of utilization that is thoroughly reasoned. This also ensures a better chance of understanding ethical implications brought forth by the algorithmic capabilities.

Ultimately, the thesis reviews a specific way in which information architecture can enrich an information environment, by establishing a semantic layer to provide significant valuable insights for different types of data-driven businesses, be it startup companies or established organizations. Through this review, it can be concluded that there is an aspect of practical value present for some data-driven businesses that implement a semantic layer to logically structure a targeted information environment.

13 Literature

5 Best Practices for Effective Personalized Product Recommendations - Mailchimp | Mailchimp. Retrieved 13 February 2021, from https://mailchimp.com/marketing-glossary/personalized-product-recommendations/?fbclid=IwAR1bl3_QYK6VMz0xLx_mHlUrhPFouw5yL2uFvVLrkVkIc2M3FqKqMPh0WgUv

Alvesson, M., & Sköldberg, K. (2000). Hermeneutics: Interpretation and Insight. In Reflexive Methodology: New Vistas for Qualitative Research. p. 66–67. London: SAGE Publications.

Arango, J. (2018). Semantic Environments and Information Architecture. Retrieved 5 May 2021, from https://jarango.com/2013/05/02/semantic-environments-and-information-architecture/?fbclid=IwAR3f8CzTjOH87cbABsTcPbnZgTkQHfR3aPvY_ObQXG1MJ1bLxii22Me5LhQ

Barth, S., & de Jong, M. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior [Ebook]. Twente NE: Elsevier Ltd. Retrieved from https://www.sciencedirect.com/science/article/pii/S0736585317302022

Bean, R., & Davenport, T. (2019). Companies Are Failing in Their Efforts to Become Data-Driven. Retrieved 23 January 2021, from https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven

Bhageshpur, K. (2019). Council Post: Data Is The New Oil - And That's A Good Thing. Retrieved 10 December 2020, from https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=4d00e5c17304

Bianchi, F., Rossiello, G., Costabello, L., Palmonari, M., & Minervini, P. (2020). Knowledge Graph Embeddings and Explainable AI. In I. Tiddi, F. Lecue & P. Hitzler, Knowledge Graphs for eXplainable AI - Foundations, Applications and Challenges (pp. 49-72). Amsterdam: IOS Press. Retrieved from https://ebooks.iospress.nl/volumearticle/54077

Bostrom, N. (2011). The Ethics of Artificial Intelligence.

Bostrom, N. (2016). Superintelligence. Oxford University Press.

Bouneffouf, D., Bouzeghoub, A., & Ganarski, A. L. (2013). Risk-Aware Recommender Systems. Neural Information Processing, 57–65. https://doi.org/10.1007/978-3-642-42054-2_8

Bryman, A., & Bell, E. (2011). Mixed methods research: combining quantitative and qualitative research. In Business Research Methods (pp. 632–645). New York: Oxford University Press.

Chee, F. (2019). EU's Landmark Data Privacy Law, GDPR, Already Providing Global Standard. Retrieved 2 May 2021, from https://www.insurancejournal.com/news/international/2019/05/23/527265.htm

Creswell, W. J. (2009). Research Design, Qualitative, Quantitative, and Mixed Methods Approaches. University of Nebraska-lincoln. Sage Publications, Inc. Third Edition.

Chmielewski, M., & Stąpor, P. (2018). Hidden information retrieval and evaluation method and tools utilising ontology reasoning applied for financial fraud analysis. In MATEC Web Conference. Retrieved from https://www.matec-conferences.org/articles/matecconf/abs/2018/69/matecconf_cscc2018_02019/matecconf_cscc2018_02019.html

Eyal, N. (2014). *Hooked: How to Build Habit-Forming Products*. New York: Penguin Books Ltd.

Foer, F. (2020). What Big Tech Wants Out of the Pandemic. Retrieved 2 May 2021, from https://www.theatlantic.com/magazine/archive/2020/07/big-tech-pandemic-power-grab/612238/

Fogg, B.J. (2009). A Behavior Model for Persuasive Design. Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09. Retrieved 26 November 2020, from: https://doi.org/10.1145/1541948.1541999

Gaming loot boxes: What happened when Belgium banned them? (2019). Retrieved 24 April 2021, from https://www.bbc.com/news/newsbeat-49674333

Glud, L. N., & Jespersen, J. L. (2008). Conceptual analysis of Kairos for Location-based mobile services. University of Oulu. Department of Information Processing Science. Series A, Research Papers (pp. 17)

Gonçalves, R. S., Horridge, M., Li, R., Liu, Y., Musen, M. A., Nyulas, C. I., ... Temple, D. (2019). Use of OWL and Semantic Web Technologies at Pinterest. Lecture Notes in Computer Science, (pp. 418–435). https://doi.org/10.1007/978-3-030-30796-7_26

Hassan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. (2013). A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. University of Lyon.

Hedden, H. (2010). The Accidental Taxonomist. Medford (NJ): Information Today. p. 69-74.

Hilbert, M. (2016). Big Data for Development: A review of Promises and Challenges, Development Policy Review. p. 135-174.

Horridge, M, et. al. (2011). A Practical Guide To Building OWL Ontologies.

How Big Tech Got Even Bigger. (2021). Retrieved 2 May 2021, from https://www.wsj.com/articles/how-big-tech-got-even-bigger-11612587632

Ibrahim, Mohammed & Yang, Yanyan & Ndzi, David & Yang, Guangguang. (2018). Ontology-Based Personalized Course Recommendation Framework. IEEE Access. 7. 5180 - 5199. 10.1109/ACCESS.2018.2889635.

Jacobson, K., & Ferris, B. (2010, August 2). The Recommendation Ontology Specification. Retrieved March 15, 2021, from http://smiy.sourceforge.net/rec/spec/recommendationontology.html

Katakam, N. (2019). How Can We 'Design' An Intelligent Recommendation Engine? Retrieved 13 February 2021, from https://uxplanet.org/how-can-we-design-an-intelligent-recommendation-engine-b9bb1db4d050

Kristine Holst, H. (2016). Kvinderne sidder på magten i hjemmet - mændene får ikke et ben til jorden. Retrieved 30 March 2021, from https://www.berlingske.dk/samfund/kvinderne-sidder-paa-magten-i-hjemmet-maendene-faar-ikke-et-ben-til-jorden

Kurzweil, R. (2005). The Singularity is Near. Viking Penguin Books. Ch. 1,2 & 7.

Lops, Pasquale & de Gemmis, Marco & Semeraro, Giovanni. (2011). Content-based Recommender Systems: State of the Art and Trends. 10.1007/978-0-387-85820-3_3.

Lops, P., de Gemmis, M., & Semeraro, G. (2010). Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook*, 73–105. https://doi.org/10.1007/978-0-387-85820-3_3

Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. *User Modeling and User-Adapted Interaction*, *29*(2), 239–249. https://doi.org/10.1007/s11257-019-09231-w

Ludewig, M., & Jannach, D. (2018). Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, *28*(4–5), 331–390. https://doi.org/10.1007/s11257-018-9209-6

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and Lucid methodology. Issues in Educational Research, 16(2). p. 193–205.

Miller, R. (2019). Big Companies Are Not Becoming Data-Driven Fast Enough. Retrieved 23 January 2021, from https://techcrunch.com/2019/02/06/big-companies-are-not-becoming-data-driven-fast-enough/

Newton, C. (2020). How COVID-19 is changing public perception of big tech companies. Retrieved 2 May 2021, from https://www.theverge.com/interface/2020/3/26/21193902/tech-backlash-covid-19-coronavirus-google-facebook-amazon

Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory, 1–25.

Number of child gamblers quadruples in just two years. (2018). Retrieved 24 April 2021, from https://www.bbc.com/news/business-46286945

Obeid, C., Lahoud, I., Khoury, H., & Champin, P. (2018). Ontology-based recommender system in higher education. In *WWW 2018*. Lyon, France: The 2018 Web Conference Companion. Retrieved from https://hal-univ-lyon1.archives-ouvertes.fr/hal-02288141

Öhgren, A., & Sandkuhl, K. (2005) Towards a methodology for ontology development in small and medium-sized enterprises. Jönköping, Sweden.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books.

Ontology editors. Retrieved 4 March 2021, from https://www.w3.org/wiki/Ontology_editors

Owens, J. (2019). The tech giants dominated the decade. But there's still time to rein them in | Jay Owens. Retrieved 2 May 2021, from https://www.theguardian.com/commentisfree/2019/dec/25/2010s-tech-giants-google-amazon-facebook-regulators

Protégé. Retrieved 4 March 2021, from https://protege.stanford.edu/

Ricci, F. (2010). Mobile Recommender Systems. Information Technology & Tourism, 12(3), 205–231. https://doi.org/10.3727/109830511x12978702284390

Rosenfeld, L., Morville, P., & Arango, J. (2015). Information Architecture: For the Web and Beyond (4th ed.). Sebastopol, California: O'Reilly Media, Incorporated.

Sanders, E B-N. & Stappers, P.J. (2008). Co-creation and the new landscapes of design. CoDesign. International Journal of cocreation in design and the arts. 4(1). 5-18.

Shah, V., & Subramanian, S. (2019). Data-Science Recommendation System using Semantic Technology. *International Journal Of Engineering And Advanced Technology*, *9*(1), 2592-2599. doi: 10.35940/ijeat.a9375.109119

Sheridan, P., Onsjö, M., Becerra, C., Jimenez, S., & Dueñas, G. (2019). An Ontology-Based Recommender System with an Application to the Star Trek Television Franchise. *Future Internet*, *11*(9), 23. doi: 10.3390/fi11090182

Sieg, A., Mobasher, B., & Burke, R. (2010). *Ontology-Based Collaborative Recommendation* [Ebook]. DePaul University, Chicago, Illinois, USA: School of Computing. Retrieved from https://www.ahusieg.com/pdf/SMB10b.pdf

Smetsers, R. (2013). Association rule mining for recommender systems. http://arno.uvt.nl/show.cgi?fid=131711

Strategy as a Wicked Problem (2014). Harvard Business Review. Retrieved 5 May 2021, from https://hbr.org/2008/05/strategy-as-a-wicked-problem#

Thanapalasingam, Thiviyan & Osborne, Francesco & Birukou, Aliaksandr & Motta, Enrico. (2018). Ontology-Based Recommendation of Editorial Products: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II. doi: 10.1007/978-3-030-00668-6_21.

Tesfaye, L. (2020). What is a Semantic Architecture and How do I Build One? Enterprise Knowledge. Retrieved 5 May 2021, from https://enterprise-knowledge.com/what-is-a-semantic-architecture-and-how-do-i-build-one/?fbclid=IwAR32nPT_-75xhXeC9wq9aSjh_ttvVeMJwTdhV5SGrJGCa7kbdfzsb-8fhRA

The Cambridge Analytica Story, Explained. Retrieved 2 May 2021, from https://www.wired.com/amp-stories/cambridge-analytica-explainer/

Theodore, G. (2020, December 9). Hermeneutics (Stanford Encyclopedia of Philosophy). Stanford.Edu. https://plato.stanford.edu/entries/hermeneutics/#HermCirc

Varzaneh, H. H., Neysiani, B. S., Ziafat, H., & Soltani, N. (2018). Recommendation Systems Based on Association Rule Mining for a Target Object by Evolutionary Algorithms. *Emerging Science Journal*, *2*(2), 100–106. https://doi.org/10.28991/esj-2018-01133

Weisbaum, H. (2018). Zuckerberg's apology tour has not done much to regain user trust. Retrieved 2 May 2021, from https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011

14 Summary

Denne afhandling gennemgår og præsenterer et praktisk afsæt i idéen om, at virksomheder burde kunne få mere værdi ud af den data de besidder ved at formidle den semantisk. Mange datadrevne virksomheder er ikke klar over, at en semantisk datastruktur åbner op for et potentiale, som kan øge værdien af deres data i form af eksempelvis en bedre brugeroplevelse; dette er tilfældet gennem logisk strukturering af relevante informationsområder, som kan argumentere for mere dybdegående viden om eksempelvis brugeres præferencer.

Den semantiske datastrukturering er blevet afprøvet i praksis, ved at have undersøgt den instrumentale værdi fra anbefalingssytemer. Anbefalingssystemet, som videreformidler indhold på baggrund af semantisk datastrukturering af information om brugere og indhold, er blevet designet som et led i et trinopdelt konceptuelt udgangspunkt, der efterfølgende skal fungere som praktisk anvendeligt på mange forskellige former for datadrevne virksomheder. I afhandlingen er værdien blevet belyst fra perspektivet af to udvalgte virksomheder, som bevidst ikke har haft samme udgangspunkt for indsamling eller brug af data. Dette valg af virksomheder har derfor været med til at give nogle forskellige perspektiver på semantisk datastrukturering samt belyse værdien fra konceptets procestrin gennem en direkte sammenligning af resultaterne virksomhederne imellem.

Konceptet har inkorporeret den formelle ontologi for at kortlægge logiske forhold mellem brudstykker af relevant data. Denne måde at redegøre for forbindelser mellem brugere og indhold videreformidler værdifuld indsigt. Dette er tilfældet, da selve ontologi-domænet kan argumenterer for uforudsete forbindelser mellem data-strenge, som gør det muligt at rationalisere viden der ikke har været eksplicit førhen. Dette kan eksempelvis være et led i at skabe mere dybdegående viden om de respektive virksomheders brugere, som en anbefalingsalgoritme netop kan agere på for at optimere brugeroplevelsen for selv samme brugere. Derudover har konceptet også haft inkorporeret teori fra adfærdsændring, for at kunne argumenterer for mere personlige anbefalinger samt at mindske risiko for en påtrængende virkning fra disse.

Afhandlingen gennemgår derudover nogle nøgleargumenter for, hvorfor et semantisk lag har nogle uforudsete værdigrundlag i forbindelse med best practices indenfor kunstig intelligensudvikling, samt hvordan et sådant lag kan være med til at give en anbefalingsalgoritme et mere menneskeligt udgangspunkt for kommunikation og videredeling af indhold. 'Menneskeligt' i den forstand, at ved sammenligning med automatiserede machine learning principper, så kan et semantisk lag redegøre for mere transparent og menneskelig argumentation.

Afhandlingen prøver at flytte teori om semantisk datastrukturering til et praktisk værdigrundlag, for på den måde at undersøge og få mere viden om den semantiske datastrukturerings praktiske anvendelse. Dette giver et afsæt, som belyser nogle vigtige indsigter om semantisk datastrukturering i forbindelse med muligheder for logisk ræsonnement i informationsarkitektur. Det belyser derudover positive effekter for datadrevne virksomheder, ved at argumentere for hvorfor disse kan få mere værdi af deres data, ved at strukturere deres data semantisk.

15 Appendix

15.1 Appendix 1 - Ontologies

The featured ontologies for OpdagDanmark and NNIT are attached to the submitted thesis. These can provide further context for understanding the semantic value that the ontologies create in both cases. There are also annotations attached for relevant classes and object properties, which support the descriptions from the thesis. Protégé is the recommended software that can be utilized to open the OWL-files.

https://protege.stanford.edu/

15.2 Appendix 2 - Literature Review

There is a total of 2157 literary pages.

Theory of Science - 327 Pages

These references are the starting point for the approach to creating scientific knowledge throughout the thesis.

Alvesson, M., & Sköldberg, K. (2000). Hermeneutics: Interpretation and Insight. In Reflexive Methodology: New Vistas for Qualitative Research. p. 66–67. London: SAGE Publications. - 4 pages

Bryman, A., & Bell, E. (2011). Mixed methods research: combining quantitative and qualitative research. In Business Research Methods (pp. 632–645). New York: Oxford University Press. - 13 pages

Creswell, W. J. (2009). Research Design, Qualitative, Quantitative, and Mixed Methods Approaches. University of Nebraska-lincoln. Sage Publications, Inc. Third Edition. - 260 pages

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and Lucid methodology. Issues in Educational Research, 16(2). p. 193–205. - 12 pages

Sanders, E B-N. & Stappers, P.J. (2008). Co-creation and the new landscapes of design. CoDesign. International Journal of cocreation in design and the arts. 4(1). 5-18. - 13 pages

Theodore, G. (2020, December 9). Hermeneutics (Stanford Encyclopedia of Philosophy). Stanford.Edu. https://plato.stanford.edu/entries/hermeneutics/#HermCirc - 25 pages

Methodology - 121 pages

State of the art review

This is the state of the art in terms of what exists already concerning semantic technologies and recommender systems.

Ibrahim, Mohammed & Yang, Yanyan & Ndzi, David & Yang, Guangguang. (2018). Ontology-Based Personalized Course Recommendation Framework. IEEE Access. 7. 5180 - 5199. 10.1109/ACCESS.2018.2889635. - 21 pages

Lops, Pasquale & de Gemmis, Marco & Semeraro, Giovanni. (2011). Content-based Recommender Systems: State of the Art and Trends. 10.1007/978-0-387-85820-3_3. - 34 pages

Obeid, C., Lahoud, I., Khoury, H., & Champin, P. (2018). Ontology-based recommender system in higher education. In *WWW 2018*. Lyon, France: The 2018 Web Conference Companion. Retrieved from https://hal-univ-lyon1.archives-ouvertes.fr/hal-02288141 - 5 pages

Shah, V., & Subramanian, S. (2019). Data-Science Recommendation System using Semantic Technology. *International Journal Of Engineering And Advanced Technology*, *9*(1), 2592-2599. doi: 10.35940/ijeat.a9375.109119 - 8 pages

Sheridan, P., Onsjö, M., Becerra, C., Jimenez, S., & Dueñas, G. (2019). An Ontology-Based Recommender System with an Application to the Star Trek Television Franchise. *Future Internet*, *11*(9), 23. doi: 10.3390/fi11090182 - 25 pages

Sieg, A., Mobasher, B., & Burke, R. (2010). *Ontology-Based Collaborative Recommendation* [Ebook]. DePaul University, Chicago, Illinois, USA: School of Computing. Retrieved from https://www.ahusieg.com/pdf/SMB10b.pdf - 12 pages

Thanapalasingam, Thiviyan & Osborne, Francesco & Birukou, Aliaksandr & Motta, Enrico. (2018). Ontology-Based Recommendation of Editorial Products: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II. doi: 10.1007/978-3-030-00668-6_21. - 16 pages

Concept - 851 pages

In this chapter, the core literature for the different steps in the concept has been accounted for.

Step 1 & 2

Bouneffouf, D., Bouzeghoub, A., & Ganarski, A. L. (2013). Risk-Aware Recommender Systems. *Neural Information Processing*, 57–65. https://doi.org/10.1007/978-3-642-42054-2_8 - 10 pages

Lops, P., de Gemmis, M., & Semeraro, G. (2010). Content-based Recommender Systems: State of the Art and Trends. *Recommender Systems Handbook*, 73–105. https://doi.org/10.1007/978-0-387-85820-3_3 - 32 pages

Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. *User Modeling and User-Adapted Interaction*, *29*(2), 239–249. https://doi.org/10.1007/s11257-019-09231-w - 10 pages

Ludewig, M., & Jannach, D. (2018). Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, *28*(4–5), 331–390. https://doi.org/10.1007/s11257-018-9209-6 - 59 pages

Ricci, F. (2010). Mobile Recommender Systems. Information Technology & Tourism, 12(3), 205–231. https://doi.org/10.3727/109830511x12978702284390 - 24 pages

Rosenfeld, L., Morville, P., & Arango, J. (2015). Information Architecture: For the Web and Beyond (4th ed.). Sebastopol, California: O'Reilly Media, Incorporated. - 451 pages

Smetsers, R. (2013). Association rule mining for recommender systems. http://arno.uvt.nl/show.cgi?fid=131711 - 74 pages

Varzaneh, H. H., Neysiani, B. S., Ziafat, H., & Soltani, N. (2018). Recommendation Systems Based on Association Rule Mining for a Target Object by Evolutionary Algorithms. *Emerging Science Journal*, *2*(2), 100–106. https://doi.org/10.28991/esj-2018-01133 - 8 pages

Step 3

Gonçalves, R. S., Horridge, M., Li, R., Liu, Y., Musen, M. A., Nyulas, C. I., ... Temple, D. (2019). Use of OWL and Semantic Web Technologies at Pinterest. Lecture Notes in Computer Science, (pp. 418–435). https://doi.org/10.1007/978-3-030-30796-7_26 - 17 pages

Hedden, H. (2010). The Accidental Taxonomist. Medford (NJ): Information Today. p. 69-74. - 5 pages

Horridge, M, et. al. (2011). A Practical Guide To Building OWL Ontologies - 104 pages

Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory, 1–25. - 25 pages

Öhgren, A., & Sandkuhl, K. (2005) Towards a methodology for ontology development in small and medium-sized enterprises. Jönköping, Sweden. - 8 pages

Step 4

Fogg, B.J. (2009). A Behavior Model for Persuasive Design. Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09. Retrieved 26 November 2020, from: https://doi.org/10.1145/1541948.1541999 - 7 pages

Glud, L. N., & Jespersen, J. L. (2008). Conceptual analysis of Kairos for Location-based mobile services. University of Oulu. Department of Information Processing Science. Series A, Research Papers (pp. 17) - 17 pages

Discussion - 858 pages

This is literature that is utilized in the discussion. A lot of what is referenced here is also part of what is considered not countable for the total literary pages. These are found in the next chapter.

Barth, S., & de Jong, M. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior [Ebook]. Twente NE: Elsevier Ltd. Retrieved from https://www.sciencedirect.com/science/article/pii/S0736585317302022 - 21 pages

Bianchi, F., Rossiello, G., Costabello, L., Palmonari, M., & Minervini, P. (2020). Knowledge Graph Embeddings and Explainable AI. In I. Tiddi, F. Lecue & P. Hitzler, Knowledge Graphs for eXplainable AI - Foundations, Applications and Challenges (pp. 49-72). Amsterdam: IOS Press. Retrieved from https://ebooks.iospress.nl/volumearticle/54077 - 23 pages

Bostrom, N. (2011). The Ethics of Artificial Intelligence. - 20 pages

Bostrom, N. (2016). Superintelligence. Oxford University Press. - 390 pages

Chmielewski, M., & Stąpor, P. (2018). Hidden information retrieval and evaluation method and tools utilising ontology reasoning applied for financial fraud analysis. In MATEC Web Conference. Retrieved from https://www.matec-conferences.org/articles/matecconf/abs/2018/69/matecconf_cscc2018_02019/matecconf_cscc2018_02019.html - 9 pages

Hassan, O., Habegger, B., Brunie, L., Bennani, N., & Damiani, E. (2013). A Discussion of Privacy Challenges in User Profiling with Big Data Techniques: The EEXCESS Use Case. University of Lyon. - 6 pages

Hilbert, M. (2016). Big Data for Development: A review of Promises and Challenges, Development Policy Review. p. 135-174. - 39 pages

Kurzweil, R. (2005). The Singularity is Near. Viking Penguin Books. Ch. 1,2 & 7. - 81 pages

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Books. - 269 pages

References for further argumentation

These references do not count in the literary pages.

5 Best Practices for Effective Personalized Product Recommendations - Mailchimp | Mailchimp. Retrieved 13 February 2021, from https://mailchimp.com/marketing-glossary/personalized-product-recommendations/?fbclid=IwAR1bl3_QYK6VMzoxLx_mHlUrhPFouw5yL2uFvVLrkVkIc2M3FqKqMPh0WgUv

Arango, J. (2018). Semantic Environments and Information Architecture. Retrieved 5 May 2021, from https://jarango.com/2013/05/02/semantic-environments-and-information-architecture/?fbclid=IwAR3f8CzTjOH87cbABsTcPbnZgTkQHfR3aPvY_ObQXG1MJ1bLxii22Me5LhQ

Bean, R., & Davenport, T. (2019). Companies Are Failing in Their Efforts to Become Data-Driven. Retrieved 23 January 2021, from https://hbr.org/2019/02/companies-are-failing-in-their-efforts-to-become-data-driven

Bhageshpur, K. (2019). Council Post: Data Is The New Oil - And That's A Good Thing. Retrieved 10 December 2020, from https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=4d00e5c17304

Chee, F. (2019). EU's Landmark Data Privacy Law, GDPR, Already Providing Global Standard. Retrieved 2 May 2021, from https://www.insurancejournal.com/news/international/2019/05/23/527265.htm

Eyal, N. (2014). *Hooked: How to Build Habit-Forming Products*. New York: Penguin Books Ltd.

Foer, F. (2020). What Big Tech Wants Out of the Pandemic. Retrieved 2 May 2021, from https://www.theatlantic.com/magazine/archive/2020/07/big-tech-pandemic-power-grab/612238/

Gaming loot boxes: What happened when Belgium banned them? (2019). Retrieved 24 April 2021, from https://www.bbc.com/news/newsbeat-49674333

How Big Tech Got Even Bigger. (2021). Retrieved 2 May 2021, from https://www.wsj.com/articles/how-big-tech-got-even-bigger-11612587632

Jacobson, K., & Ferris, B. (2010, August 2). The Recommendation Ontology Specification. Retrieved March 15, 2021, from http://smiy.sourceforge.net/rec/spec/recommendationontology.html

Miller, R. (2019). Big Companies Are Not Becoming Data-Driven Fast Enough. Retrieved 23 January 2021, from https://techcrunch.com/2019/02/06/big-companies-are-not-becoming-data-driven-fast-enough/

Katakam, N. (2019). How Can We 'Design' An Intelligent Recommendation Engine? Retrieved 13 February 2021, from https://uxplanet.org/how-can-we-design-an-intelligent-recommendation-engine-b9bb1db4d050

Kristine Holst, H. (2016). Kvinderne sidder på magten i hjemmet - mændene får ikke et ben til jorden. Retrieved 30 March 2021, from https://www.berlingske.dk/samfund/kvinderne-sidder-paa-magten-i-hjemmet-maendene-faar-ikke-et-ben-til-jorden

Newton, C. (2020). How COVID-19 is changing public perception of big tech companies. Retrieved 2 May 2021, from https://www.theverge.com/interface/2020/3/26/21193902/tech-backlash-covid-19-coronavirus-google-facebook-amazon

Number of child gamblers quadruples in just two years. (2018). Retrieved 24 April 2021, from https://www.bbc.com/news/business-46286945

Ontology editors. Retrieved 4 March 2021, from https://www.w3.org/wiki/Ontology_editors

Owens, J. (2019). The tech giants dominated the decade. But there's still time to rein them in | Jay Owens. Retrieved 2 May 2021, from https://www.theguardian.com/commentisfree/2019/dec/25/2010s-tech-giants-google-amazon-facebook-regulators

Protégé. Retrieved 4 March 2021, from https://protege.stanford.edu/

Strategy as a Wicked Problem (2014). Harvard Business Review. Retrieved 5 May 2021, from https://hbr.org/2008/05/strategy-as-a-wicked-problem#

Tesfaye, L. (2020). What is a Semantic Architecture and How do I Build One? Enterprise Knowledge. Retrieved 5 May 2021, from https://enterprise-knowledge.com/what-is-a-semantic-architecture-and-how-do-i-build-one/?fbclid=IwAR32nPT_-75xhXeC9wq9aSjh_ttvVeMJwTdhV5SGrJGCa7kbdfzsb-8fhRA

The Cambridge Analytica Story, Explained. Retrieved 2 May 2021, from https://www.wired.com/amp-stories/cambridge-analytica-explainer/

Weisbaum, H. (2018). Zuckerberg's apology tour has not done much to regain user trust. Retrieved 2 May 2021, from https://www.nbcnews.com/business/consumer/trust-facebook-has-dropped-51-percent-cambridge-analytica-scandal-n867011

15.3 Appendix 3 - Literature Acceptance

David Jakobsen

fr 07-05, 16:50

Undskyld, det er denne som er godkendt! J

...

Fra: Joachim August Overgaard <joverg19@student.aau.dk>
Sendt: 7. maj 2021 16:48
Til: David Jakobsen <davker@hum.aau.dk>
Cc: Nicklas Thorup Poulsen <npouls19@student.aau.dk>
Emne: Sv: Litteratur til acceptance

Hej David,

Vi har modificeret den og tilføjet side tal for hvert stykke litteratur og for de overordnede overskrifter.

Fra: David Jakobsen
Sendt: 7. maj 2021 16:03:02
Til: Joachim August Overgaard
Cc: Nicklas Thorup Poulsen
Emne: SV: Litteratur til acceptance

Kære Joachim og Nicklas

Jeg har lige brug for at i, efter hvert punkt, skriver sidetal så jeg kan se hvor meget i opgiver.
Mvh
David

Fra: Joachim August Overgaard <joverg19@student.aau.dk>
Sendt: 7. maj 2021 15:47
Til: David Jakobsen <davker@hum.aau.dk>
Cc: Nicklas Thorup Poulsen <npouls19@student.aau.dk>
Emne: Litteratur til acceptance

Hej David,

Vi har fået samlet vores litteratur for specialet.

Derfor sender vi det dig her i mailen, for at få det accepteret.

Skriv endelig hvis du har nogle spørgsmål eller noget input til det.

Mvh,
Joachim og Nicklas

David Jakobsen

fr 07-05, 16:49

Det er fint. Den er godkendt!

...