
A COMPUTER MODEL OF INFORMATION BEHAVIOR TO STUDY INFORMATION SECURITY PROFESSIONALS

Paulo Hideo Ohtoshi (1), Claudio Gottschalg-Duque (2)

(1) School of Information, University of California, Berkeley, USA and Faculty of Information Science, University of Brasilia, Brazil, ohtoshi@gmail.com, ohtoshi@unb.br (2) Faculty of Information Science, University of Brasilia, klaus@unb.br

Abstract

This paper proposes a computer model of information behavior to study information security professionals and an architecture, which mimics the way our brain learns new concepts to simulate this behavior computationally. Used to represent and describe any domain of knowledge, ontologies may be used to study human information behavior and show some of the concepts and relationships involved in this field of knowledge. A deep knowledge of the core concepts underpinning this field can provide us with a solid basis for constructing a model. Computer-programming tools can be used not only to capture the ideas that make up this field of Know-

ledge but can also simulate the human information behavior. The use of computers also allows us to crawl data over the Internet and process large amounts of them in order to find patterns with some specific characteristics. In the paper, we also present the current state of this research and the challenges of the model.

Keywords: Information behavior; ontologies; artificial intelligence; machine learning; predictive models; psychology; cognitive science

1 Introduction

The Web has changed how people communicate with each other, how information is disseminated and retrieved, and how business is conducted (Antoniou & Harmelen, 2012). It has not only affected the way we communicate with other people, but also the way information is dealt with, in order to perform our daily tasks and activities. All of this is to show that information behavior will always change over time and be an object of permanent discussion. Therefore, new and dynamic models are needed to describe and understand these changes.

Wilson (1981) refers to some questions of information behavior as intractable issues. He refers to the information need as a subjective experience which occurs only in the mind of a person and, consequently, not accessible to an observer. According to Dervin (1983), information need is a cognitive gap. Information need is a state which arises when someone is executing a task or an activity that may require information. Information need is one of the concepts that have caused some difficulties in information behavior studies. Due to the non-procedural nature of some of the aspects of this field of knowledge, they cannot be adequately expressed by using block diagrams and flowcharts. Based on the assumption that ontologies may be used to describe any phenomenon, this research uses them to describe the human information behavior. The conceptual model proposed describes the whole cycle of the

human information behavior: need, seeking and use of information. The research question to be answered by this research is: Is it possible to create an integrated model to study the information behavior (need, seeking and use) of professionals in their workplace and ordinary people in the daily life?

Models are of great value in the development of theory. Models are more useful at the description and prediction stages of understanding a phenomenon. When we develop an explanation for a phenomenon, we can properly say we have a theory (Bates, 2005). In general terms, developing a theory of something is developing a view, a description, or a way of looking at it. A good description not only depicts something but also explains and fosters understanding of it (Buckland, 1991).

This paper presents an ontology-based model of information behavior and proposes a modular and incremental architecture to simulate this behavior using computers. The ontology mimics the cognitive structure of our brain and the architecture suggested mimics the way our brain learns new concepts to construct knowledge (Ausubel, 2011). The model uses ideas from different fields for studying the information behavior of information security professionals. These professionals are responsible for keeping the confidentiality, integrity and availability of organizational information. The model uses concepts from Information Behavior, Ontologies, Semantic Web, Artificial Intelligence, Machine Learning, Psychology, and Cognitive Science. It was proposed to study information security profession-

als, but it may be adapted to study both professionals in their work environment and ordinary people in their daily lives and out of the work environment. Scalability, reusability and knowledge sharing are some of the features of ontologies which allow us to adapt it easily for studying different categories of people.

The architecture presented uses ontologies dynamically constructed, allowing the computer to acquire data from the environment and to add autonomously new concepts to the previous structure. In other words, the machine may start to learn new concepts and improve their knowledge of a certain domain of knowledge.

The architecture was divided into three modules and four steps of development:

- Module 1 - Processing Module (Ontology):
 - a) Conceptual Model (Diagram) – Human-Readable - Step 1
 - b) Computer Model (Computer Language) – Machine-Readable - Step 2
- Module 2 - Data Storage Module (Database) – Step 3
- Module 3- Data Acquisition Module – Step 4

For the purpose of this research, and due to the time constraints, only the Conceptual Model (diagram) of the ontology, which corresponds to the Step 1 of Module 1 will be constructed and validated. The Conceptual Model is a graphical representation (human-readable) of the ontology. Even though the whole architecture cannot be implemented, we decided to publish it because it does not depend on the complete ontology. Like our brain, the architecture proposed can be used to construct ontologies for any domain of knowledge. In addition, each module of the architecture may be constructed independently.

In the architecture description, each module will be discussed and presented. With the technology available, the construction of Modules 2 and 3 are completely feasible. For Module 1, there are technologies available to construct dynamically any ontology. The challenge of Module 1 is not the ontology construction, but some key issues of the domain represented by the conceptual model.

The ontology model of information behavior, like any other ontology, is a scalable structure, whose development may be incremental. The machine can be taught, implementing the initial conceptual model as a static structure, and then, let the machine learn by itself, adding up dynamically and automatically new concepts (classes, instances, and properties) the system gets from the environment.

Ontologies allow us to make inferences based on deductive logic reasoning. The more information we have about people the better inferences we can make about them. Based on the ideas of predictive models, we

believe that the machine may find patterns of behavior on the data acquired. The quantity of information we have about a person affects the inferential capability.

The Model of Information Behavior will be presented in two versions: A Conceptual Model (diagram), a graphical human-readable representation, which is being validated using empirical data, and a Computer Model (computer program), which will be a formal machine-readable representation (programming language) and based on the conceptual model.

One of the most challenging problems to study small groups of people, such as information security professionals, is the generalization. Due to the need to validate the Conceptual Model (diagram) and the lack of large amounts of data, a qualitative approach was applied. The qualitative method was ideal to capture some behavioral aspects described in detail by the people interviewed.

In the second step, the Computer Model development will allow the use of quantitative methods in order to analyze large amounts of data. Like the ideas employed in predictive models used in machine learning, we will collect many data from ordinary people trying to find these patterns. Kelleher, Namee, and D'Arcy (2015) define machine learning as an automated process that extracts patterns from data.

This is a key issue of this proposal, to realize that information security professionals are a subcategory (subclass) of ordinary people category (class). Using Venn diagram notation, where P is the set of professionals and O the set of ordinary people: P is a subset of O.

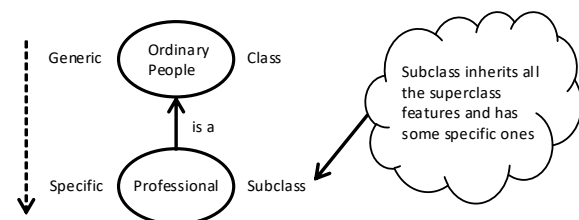


Figure 1. Class and Subclass Relationship

Figure 1 illustrates the class and subclass relationship and Figure 2 illustrates the set and subset relationship. In summary, they share or inherit most of the ordinary people's characteristics and have some specific ones. In other words, there isn't enough data about information security professionals to generalize some results but based on the assumption that they are a subclass for ordinary people, we will use ordinary people's data to construct our Computer Model (computer program) of the ontology and gradually add some specific features about information security professionals to it.

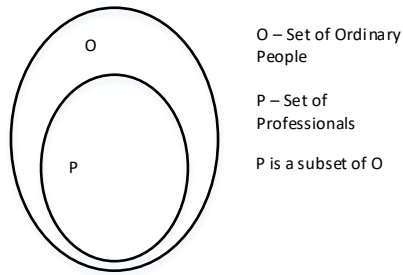


Figure 2. Set and Subset Relationship

A prototype of a car is an example of this approach. A prototype of an ordinary car may be constructed to start studying its basic behavior and gradually some specific features of a sportier car, such as a different design and more powerful motor, may be added to analyze the influence of these specific features. Similarly, a model may be created with ordinary people's features (general) and gradually professional's features (specific) may be added to it.

When talking about ordinary people, we are talking about a population of 7.6 billion people around the world, according to the UN Department of Economic and Social Affairs (2017). Among them, more than 2.5 billion use a social media program, according to Statista (2018). With the advance in computer technology, we currently can store and process large amounts of data, as well as to access it from physically distant locations over a computer network (Alpaydin, 2014). Information about ordinary people can be obtained from services like Facebook and Twitter and about professionals using services like LinkedIn or similar services. Large amounts of data become useful only when they are analyzed and turned into information, which can be used, for example, to make predictions.

In the literature about information behavior, there are different models to study ordinary people and professionals as if they were disjoint classes. Studying them separately, some important and common features and relations, which influence their behavior might not be observed.

2 Ontologies

An ontology, as defined by Gruber (1993), is a "formal, explicit specification of a shared conceptualization of a domain of interest". Typically, we represent an ontology as a hierarchical data structure containing all the relevant entities, their relationships and the rules within a domain (Leung, 2011, p.99). A 5-tuple based structure (Maedeché, 2002) is a commonly used formal description to describe the concepts and relationships in a domain; the 5-tuple core ontology structure is defined as:

$$O = (C, R, H, rel, A)$$

Where:

- C is the concepts set.
- R is relation set.
- H stands for the hierarchy of concepts.
- *rel* stands for relations among concepts
- A is the axioms set

rel is defined as a set of 3-tuple relations: $rel = (s, r, o)$, standing for the relationship of subject-relation-object, where *s* is the subject element from C, *r* is the relation element from R, and *o* is the object element from C.

3 Conceptual Model

The first step comprises the development and validation of the Conceptual Model. We based its construction on the theories and models found in the literature about information behavior studies. The concepts and basic ideas of the models proposed by Wilson (1981, 1999), Leckie, Pettigrew and Sylvain (1996), and Choo (2000) constitute the basis for the construction of this model. It was constructed based on the theoretical models and not on the observation of the phenomena. Knowledge engineering (Gašević, Djurić, and Devedžić, 2009) uses some methods such as interviews, questionnaires, observation of task performance, protocol analysis (asking the domain expert to "think aloud" while performing a task) as a form of acquisition of knowledge for constructing ontologies.

Some of these techniques are being used not to build the model, but to validate it. Data are being collected using questionnaires, interviews and document analysis from the population studied to validate empirically the model. As soon as the empirical tests are finished, some changes will be made to reflect the results achieved from the data collected. Even before finishing the analysis of the results, it is possible to realize that some intervening factors described in the Conceptual Model that affect people, such as "*Stress/Cope with Theory*", mentioned by Wilson (1999), also affect information security professionals, the population studied in this research.

One of the benefits of ontologies is the scalability, which can be seen in a graphical representation of the model. The main level shows a holistic view of the entire phenomenon, as illustrated by Figure 3 (Appendix), the Conceptual Model. Another level may be used to show some details of it. This kind of approach allows focusing on certain details without losing the general idea. The human brain performs better if new concepts are added in a hierarchical fashion, from general concepts to more specific and detailed ones (Novak, 1985).

Information security professionals are a heterogeneous population made up of experts from different fields of knowledge. Working in an information-driven environment, mainly dominated by computer science pro-

professionals, they must deal with a lot of information. Sometimes they feel overwhelmed by the amount of information available as well as by the myriad of technologies involved in their daily activities. Some of them work with short-term and daily activities such as network security management, while others deal with long-term activities, such as cryptographic hardware design. In each case, they may have to look for information available from different sources: printed or digital, formal or informal, people, books and papers. The conceptual model described in this paper represents all these characteristics.

The central and main concept of the model is the “*IS Professional*”, the user of information, which represents information security category of professionals and is a subclass of “*Person*”, which has a “*Demographic Characteristics*” and “*Types of Personality*”. As mentioned before, due to this class and subclass relationship between ordinary people and professionals, professionals inherit all the characteristics of people. This feature allows us to build a model to study both categories at the same time. “*Demographic Characteristics*” such as age, gender and social-economic features define personal features. The “*Types of Personality*” concept use the typology proposed by Myers-Briggs (Myers & Myers, 1980) and based on Carl Jung’s theories. The essence of the theory is that much seemingly random variation in the behavior is actually quite orderly and consistent, being due to basic differences in the way individuals prefer to use their perception and judgment. Myers-Briggs defined 16 different types of personality, which are a combination of four dichotomies: extraversion *versus* introversion; sensing *versus* intuition; thinking *versus* feeling; and judging *versus* perceiving. These features may influence and affect many aspects of human information behavior, hence justifying the importance of correct categorization.

As a general model, it can be used to study other categories of people. This model uses a concept that is like a component used in object-oriented programming or a table in relational databases. For example, a well-standardized component or table used to describe the “*IS Professional*” can be replaced by a component or table which represents another category, such a lawyer or a doctor.

One important contribution of this model is a better characterization of some important issues. One example: In the model proposed by Leckie *et al.* (1996), the *status in the organization*, *years of experience*, the *field of specialization* are denominated “*Characteristics of Information Needs*” that shape the information need. In this model, we designate these aspects as “*Professionals Characteristics*” and not “*Characteristics of Information Needs*”. This subtle difference is essential in the characterization and understanding of human behavior,

especially when defining classes, subclasses and inheritance issues.

“*IS Professional*” may play many “*Roles*” such as cryptographic hardware developers, cryptographic protocols and algorithm designers and cryptanalysts, network security managers and information security managers. These “*Roles*” require the execution of some “*Tasks/Activities*” such as a protocol design, development of security policies, and application of security measures. These “*Tasks/Activities*” in turn may require some “*Information*”. The lack/gap of this “*Information*” generates “*Information Needs*”. Dervin described information need a cognitive gap (Dervin, 1983) and Belkin, as an anomalous state (Belkin, 1980) and Taylor, as levels (Taylor, 1968).

In order to fill this gap or the lack of information, the “*IS Professional*” may start “*Information Seeking*” to look for “*Information*”. To seek information, the “*IS Professional*” may use some strategy to search for it. The “*Seeking Strategy*” uses the model proposed by Ellis (1989), one of the most widely cited in the information behavior literature, along with Kuhlthau’s (1993) model, and suitable to describe the information seeking in digital environments, such as the Internet. The Ellis’ model, which is not set out as a diagrammatic model, consists of eight activities: 1) Starting; 2) Chaining; 3) Browsing; 4) Differentiating; 5) Monitoring; 6) Extracting; and 7) Verifying. There another important model, which describe the search for information proposed by the Kuhlthau (1993).

The process “*Information Seeking*” seeks “*Information*” in “*Information Sources*”. These sources may be formal (manual or computer-based) or informal (people). The preference for accessible sources seems to conform to Zipf’s Law or Principle of Least Effort (Zipf, 1965). “*Intervening Factors*”, such as affective, cognitive or situational factors, and the activating mechanism suggested by Wilson (1999) affect the need, seeking and use of information.

The final stage of the model “*Information Use*” describes the individual actions on the information found. Taylor (1996) proposes that one or more of just eight categories may describe the ways people use information. In this model, we propose five categories or “*Types of Use*”: problem-solving, learning, storage, decision-making, and information exchange.

4 Empirical Results for the Validation Process

As an ongoing research, we are currently collecting data from a population of information security professionals using a sample of 59 respondents to empirically verify and validate the model. We have selected 10 of these respondents for an interview, which have allowed us to get detailed answers and long narratives. This qualitative approach has been suitable to study the behavior of this population, which is relatively small,

compared to scientists and engineers, but not big enough to generalize.

We are now working on the data collected empirically, especially the data derived from the interviews. There are some interesting results, some expected and others completely unexpected, but all of them seem to present some patterns of behavior and association. Despite presenting some apparent relations, we cannot state or prove these relations using only a qualitative model. This is the reason for proposing a computer model to extract large amounts of data to validate them.

From the data collected, we can realize that the professionals of the field come from different fields and have distinct backgrounds. This is quite different from fields such as astronomy, physics, and chemistry, where the categories of professionals are less heterogeneous. In some fields of knowledge, for example, medicine, we can say that a doctor studied medicine, but not all the people who studied medicine are doctors (Figure 4).

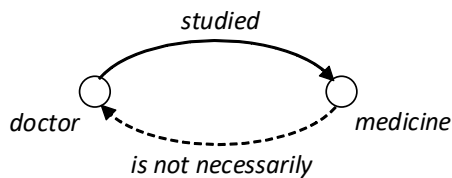


Figure 4. *Non-invertible relations*

The same reasoning does not apply to the information security professionals. It is not possible to say that an information security professional has necessarily a degree in computer science.

Four subgroups were identified and subdivided according to the roles played and academic background: cryptographic hardware designers (electrical engineering), cryptographic protocol and algorithms designers (mathematics and statistics), network security managers (computer science, electrical engineering), and information security managers (multiple fields). Each subgroup must cope with different tasks, which in turn, present distinct characteristics and levels of complexity (Byström and Jarvelin, 1992).

Some factors may affect, stimulate or inhibit behavior, acting positively or negatively on human's information behavior, such as the "*Stress/Cope with Theory*" or "*Information Avoidance*", which affect people and prevent them from looking for information that may produce negative feeling. The same also occurs with information security professionals. Again, "*Stress/Cope with Theory*", may affect people in the studies related to health and certain diseases. People suffering from incurable diseases may avoid seeking information that may produce some stress caused by negative feelings.

There is also another factor mentioned by Wilson (1999) in the literature, which is the "*Principle of Least Effort*" (Zipf, 1965). According to the theory, individuals adopt a course of action that will expend the probable least of their work or effort and tend to use the most convenient search method in information seeking. The seeking of information stops as soon as minimally enough information has been found. The results demonstrate that the effort required to find information has changed over time. Modern technologies and easy access to a computer to look for information and search engines capable of ranking the best occurrences of such information have reduced the amount of effort necessary to find information. However, technology does not affect the underlying principle that applies when the non-trivial amount of effort is needed. No matter how efficient the systems become, people will ordinarily choose the options that expend the least effort.

Several empirical studies have found that, as knowledge of a source, its potential contents and capabilities increases, the use of that source tends to increase; that is, humans tend to return to the sources that they have used in the past in strong preference to try out new sources of information. Leckie (2005) describes this phenomenon as "*Awareness of Information*". The length of experience influences "*Awareness of Information*". The more experienced the professionals the more familiar they are with the sources of information.

Some social aspects may also influence positively information seeking. In some situations, people avoid showing unawareness about something they think they should have known. Some experts avoid asking colleagues for information about something they do not know and go to look for it in another opportunity.

Another important factor, which the literature constantly mentions is "*Information Overload*". With a lot of information about information security flaws, the excess of information constantly overwhelms these professionals. One of the professionals interviewed, a security incident response expert said that the use of an internet browser such as Google has helped them to overcome or reduce the effects of this problem. According to him, the ranking capability of Google has helped them to cope with large amounts of information. The top 10 items in the ranked search solve most of the problems. He said he did not even go to the other pages. Other experts, on the other hand, have demonstrated that, sometimes they must expend much more effort than that. Especially the tasks that require complex information, which are not always easily available, may demand more effort.

Although the empirical results have shown some outstanding insights, an extensive research must be conducted using a computer model to get reliable conclusions.

5 The Computer Model – A Future Work

Only the Conceptual Model (Step 1) has been created and tested. The remaining three steps of the 4-step architecture, which are out of the scope this report, will be implemented after concluding the first one. A key issue presented in this proposal is to implement the Conceptual Model (graphical representation) of the ontology and then create the Computer Model (formal computer language representation) using an ontology editor. This initial static ontology will act as a minimum knowledge required for the machine to start learning. The system will add the other concepts gradually and dynamically, as the machine starts capturing new data.

After validating the Conceptual Model (Step 1) empirically, the second step of this research will be to construct the Computer Model (Step 2). As pointed out, the ontology model may be gradually developed. Its development may start with just a few classes and new ones may gradually be inserted, verifying, in each step, the model consistency, using the *reasoner*. In computer ontology, a *reasoner* is a tool used to check the logic of the model in order to identify some inconsistencies.

The idea of ontologies emerged in applied artificial intelligence for sharing knowledge (Gašević, Djurić, & Devedžić, 2009). An ontology resembles the cognitive structure of our brain. We learn by connecting new concepts to the previous structure, as illustrated in Figure 5. Each of these connections creates meaning and generates semantic networks. Our brain not only connects new concepts but also constantly rearranges these structures in order to create knowledge (Ausubel, 2011).

Constructing ontologies structures using editors such as *Protégé*¹, produce static structures. Whenever another class (concept) or an instance (individual) must be added, the ontology must be opened to do so. The use of a computer programming language to construct ontologies dynamically allows us to modify them creating non-static ontology structures. We can dynamically, add, remove elements and change the ontology structure. This is like how our brain deals with information while constructing knowledge (Ausubel, 2011). If the system starts adding up news concepts to itself, the machine will be emulating the learning process.

The idea is to create the Computer Model of the ontology using the *Protégé* editor, a static structure, and save as an Owl (Web Ontology Language) file. Using a computer language, such as Python, this file ontology may be opened, and dynamically new classes, subclasses, and properties may be created. This initial static structure will function as a previous knowledge from which the computer will start to learn (Figure 5).

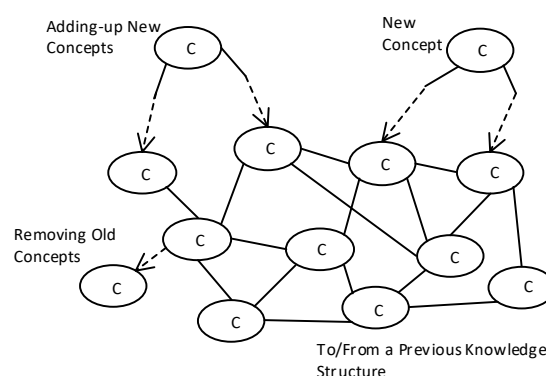


Figure 5. A graph representing the development of the ontology

Allen Newell (1982) introduced the notion of levels of knowledge and their representation. He defined three levels: implementation, logical and knowledge. On the implementation levels, he previewed structures to hold the domain knowledge, which corresponds to the initial static structure of the ontology, as well as those deployed to hold the current problem description or the working memory, which corresponds to the populated ontology used to process the simulation. The long-term memory corresponds to the database used to store data about people.

Computer tools also allow us to populate this ontology with data from specific people (instances). The use of computer programs allows us to process a huge amount of data. The capability of learning of the model may allow the machine to cope with complex systems. A complex system is a system made up of a set of many simple relations.

Starting with just a few classes (concepts), or with a less complex system, to check and test the machine, we can verify its functioning and gradually add new classes (concepts) and instances (individuals), in an incremental basis. We may supervise the model and decide which of new attributes, relations, and concepts the machine may add to itself. It is important to point out that we might not accept all the logical relations. Something might be logical, but wrong. Like any human being, a machine may also learn and make mistakes and get wrong ideas or interpretations from the environment. Therefore, we must inspect each new concept before letting the system incorporate it.

Artificial intelligence is the range of technologies that allow computer systems to perform complex functions mirroring the workings of the human mind. In AI, knowledge storing is the process of putting knowledge, encoded in a suitable format, into computer memory. Knowledge retrieval is the inverse process, finding knowledge when it is needed (Gašević, Djurić, & Devedžić, 2009). The definition presented by these

authors corresponds to what the architecture proposed in this paper does.

6 Some Important Concepts about Ontologies

Some concepts of ontologies used in this paper must be outlined. The basic (semantic) unit of ontologies is the concept. Alone, a concept may not convey too much meaning. As you connect this isolated concept to another concept by a relation, you start to create (construct) and convey more meaning. The more connections you create, the more meaningful it becomes. You have a semantic network.

Figure 6 shows a semantic network: a network of concepts connected to create meaning. Like atoms connected to create molecules and molecules connected to create more complex structures and substances, we may connect two concepts by an arc to form a proposition and connect many propositions to form a semantic network. The more connections you make the more meaningful the structure becomes and the more knowledge we add to the previous structures. This structure is like the human cognitive structure (Ausubel, 2011).

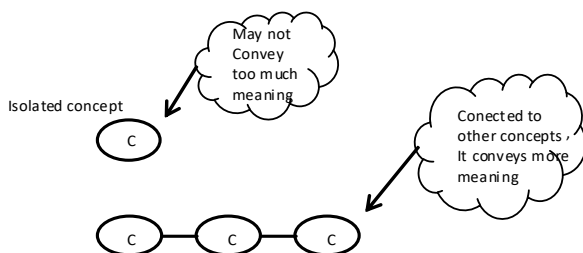


Figure 6. Constructing a semantic network

For example, the concept “Billy” isolated does not convey too much meaning besides the idea that it is a name of something, but when connected to another concept like a “dog”, now it is possible to infer that Billy is a dog and not a man and has four legs.

7 The System Architecture

Three modules make up the system architecture (Figure 7). They could be constructed in C++, Java or Python. The Python was selected to construct the model for many reasons. The most important one is the availability of scientific libraries such as *NumPy*, *MatPlotLib* and *SciPy* (Muller and Guido, 2017).

Python also provides a series of open source libraries for implementing machine learning algorithms (Raschka, 2015). There is a library to construct ontologies developed by the University of Paris, Sorbonne.

This Library, called *Owlready*², allows the creation of classes, subclasses, and instances. The documents of this library present some code based on a pizza ontology created by the University of Manchester.

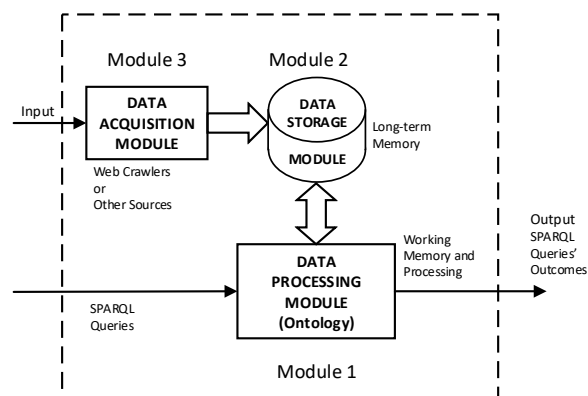


Figure 7. System architecture

7.1 Module 1 - Data Processing Module

Composed by the Computer Model of the ontology, the Data Processing Module is responsible for receiving the data from the Data Storage Module. To mimic the working memory, the system may delete the data of the ontology after using it to simulate a person. The outputs of the system are the results of the SPARQL³ queries created by the system’s users. SPARQL is a semantic query language for databases used to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

7.2 Module 2 - Data Storage Module

To store the data collected from the Internet by the web crawlers, a MySQL relational database will be constructed using Python libraries. MySQL is one of the most used open-source relational databases, which allows import and export data from/to other database formats. This module will contain tables to store data about people, such as demographic features, types of personality, professional features. The data may also be fed up directly instead of being captured from the Internet.

7.3 Module 3 - Data Acquisition Module

To crawl the data from the Internet, Russel (2014) suggests some useful techniques. For example, the author introduces techniques for mining data at LinkedIn⁴, a social networking site focused on professional and business, using an Application Program Interface (API). He also provides techniques for crawling data from other social networks, such as Facebook and Twitter, and points out the sensitive nature of LinkedIn’s data. People who join LinkedIn are primarily interested in the business opportunities that it provides as opposed to arbitrary socializing and will necessarily be providing sensitive details about business relationships, job histories, and more.

8 The Challenges of the Model

Some key points of this model must be outlined. The model may be applied to study both ordinary people

and other categories of professionals. The theory states that there are relationships between types of personality and decision and choices someone makes. In general, psychologists use questionnaires to identify personality traits. Based on the work of Adali & Golbeck (2014), the model will predict personality traits using the data collected from Facebook and Twitter. One important issue that must be addressed by the model is the difference between group characteristics and individual characteristics. In general, the model might reflect patterns of behavior. Individual characteristics of specific people will only be interpreted by the model if a determined set of characteristics fits into a certain pattern. For example, a set of personality traits must correspond to one of the 16 types proposed by Myers & Myers (1980). Another question, which will be addressed during its development, is the criteria used to accept new relationships and concepts as regularities. One possible solution for this problem is to look for patterns using machine learning algorithms. Each regularity or proposition is a pattern. For example, a dog has four legs is a regularity because most dogs meet this pattern.

9 Conclusions

Today, with the availability of large amounts of data, the challenge is not to get data, but to process them adequately and explore these resources in a meaningful way. Trying to foster innovative research, we have used ideas from seven fields of knowledge, which have strongly evolved in recent years, to construct a model and propose a different way of addressing information behavior. The model presented in this paper might allow a better understanding of human information behavior, and, by describing such phenomenon and its concepts and relationships, it might contribute to the theoretical framework for the field of information behavior studies in information science.

The capability of predicting human information behavior may allow us to anticipate facts and prepare ourselves for facing future events. Feeding the model with data collected from different sources, we may use it for different purposes. One of the noteworthy benefits associated with the capability of simulating computationally the human information behavior is to improve information services to meet the needs of information users. The primary goal of this model is to allow the understanding of the human information behavior. However, due to its adaptability, several fields of knowledge may benefit from it. The fact that it mimics the user's information behavior also makes it suitable for the study of other categories of people. In marketing, redirecting or focusing on marketing strategies. In criminal and cyber intelligence fields (McCue, 2015), by providing means to systematically and methodologically study criminal activities and prevent someone from committing a crime or acts of terror.

We will present the progress of this ongoing research and its results in a series of papers.

Notes

- (1) <http://protege.stanford.edu>
- (2) <https://pypi.python.org/pypi/Owllready>
- (3) SPARQL - SPARQL Protocol and RDF Query Language
- (4) <https://www.linkedin.com/>

Translated by the authors

References

- Adali, S.; Golbeck, J. (2014). Predicting personality with social behavior: a comparative study. // *Social Network Analysis and Mining* 4:1 (2014).
- Antoniou, G.; Harmelen, F. V. (2012). *A semantic web primer*. Cambridge, MA: MIT Press, 2012.
- Alpaydin, E. (2014). *Introduction to machine learning*. Cambridge, MA: MIT Press, 2014.
- Ausubel, D. P. (2011). *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Springer, 2011.
- Bates, M. (2005). An introduction to metatheories, theories, and models. // Fische, K.; Erdelez, S.; Mckechnie, L. (Orgs.). *Theories of information behavior*. Medford: Information Today, 1-24.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. // *Canadian Journal of Information Science* 5 (1980) 133-143.
- Buckland, M. K. (1991). *Information and information systems*. New York, NY: Praeger Publishers, 1991.
- Byström, K., Järvelin, K. (1995). Task complexity affects information seeking and use. // *Information Processing and Management* 31 (1995) 191-213.
- Choo, C. W., Detlor, B., & Turnbull, D. (2000). *Web work: Information seeking and knowledge work on the World Wide Web*. Dordrecht: Kluwer Academic, 2000.
- Choo, C. W. (2006). *The Knowing organization: how organizations use information to construct meaning, create knowledge, and make decisions*. New York: Oxford University Press, 2006.
- Dervin, B. (1983). An overview of sense-making research: Concepts, methods and results to date. // *International Communications Association Annual Meeting*, Dallas, Texas, USA, 1993.
- Ellis, D. (1989). A behavioral approach to information retrieval system design. // *Journal of Documentation* 45:3 (1989) 171-212.
- Gašević, D., Djurić, D.; Devedžić, V. (2009). *Model driven engineering and ontology development*. Berlin: Springer, 2009.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontologies. *Knowledge Acquisition* 5:2 (1993) 199-220.
- Kelleher, J. D.; Namee, B. M.; D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Massachusetts: The MIT Press, 2015.
- Kuhlthau, C. C. (1993). *Seeking meaning: a process approach to library and information services*. Norwood, NJ: Abex Publishing, 1993.
- Leckie, G. J.; Pettigrew, K. E.; Sylvain, C. (1996). *Modelling the Information Seeking of Professionals: A General Model Derived*

- from Research on Engineers, Health Care Professionals, and Lawyers. // *The Library Quarterly* 66:2 (1996) 161-193.
- Leckie, G. J. (2005). General model of the information seeking of professionals. // Fisher, K. E.; Erdelez, S.; McKechnie, L. E. F. (Eds.). *Theories of information behavior*. Medford, N.J.: Published for the American Society for Information Science and Technology by Information Today, 2005. 158-163.
- Leung, H. (2011). *Service intelligence and service science: evolutionary technologies and challenges*. Hershey, PA: IGI Global, 2011.
- Maedeche, A. D. (2002). *Ontology Learning for the Semantic Web*. New York: Springer-Verlag, 2002.
- McCue, C. (2015). *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*. Burlington, MA: Butterworth-Heinemann / Elsevier, 2015.
- Myers, I. B., & Myers, P. B. (1980). *Gifts differing: Understanding personality type*. Mountain View, CA: CPP, 1980.
- Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python: A guide for data scientists*. Sebastopol, CA: O'Reilly, 2017
- Newell, A. (1982). *The knowledge level*. Pittsburgh, PA: Design Research Center, Carnegie-Mellon University, 1982.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press, 1984.
- Raschka, S. (2015). *Python Machine Learning*. Birmingham: Packt Publishing Limited, 2015.
- Russel, M. A. (2014). *Mining the social web: data mining Facebook, Twitter, LinkedIn, Google, Github, and more*. Sebastopol, CA: O'Reilly, 2014.
- Statista (2018). Number of social media users worldwide from 2010 to 2021. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Taylor, R. S. (1968). Question-Negotiation and Information Seeking in Libraries. // *College & Research Libraries* 29:3 (1968) 178-194.
- Taylor, R. S. (1996). Information use environments. // E. Auster & C. W. Choo (Eds.), *Managing Information for the competitive edge*. New York: Neal-Schuman, 1996. 93-135.
- UN Department of Economic and Social Affairs – Population Division (2017). *World Population*.
<http://www.un.org/en/development/desa/population/>
- Wilson, T. D. (1981). On user studies and information needs. // *Journal of documentation* 37:1 (1981) 658-670.
- Wilson, T. D. (1999). Models in information behavior research. // *Journal of documentation* 55:3 (1999) 249-270.
- Zipf, G. K. (1965). *Human behavior and the principle of least effort. An introduction to human ecology*. Hafner Publishing Company: New York & London, 1965.

Copyright: © 2018. Ohtoshi e Gottschalg-Duque. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Appendix

Figure 3

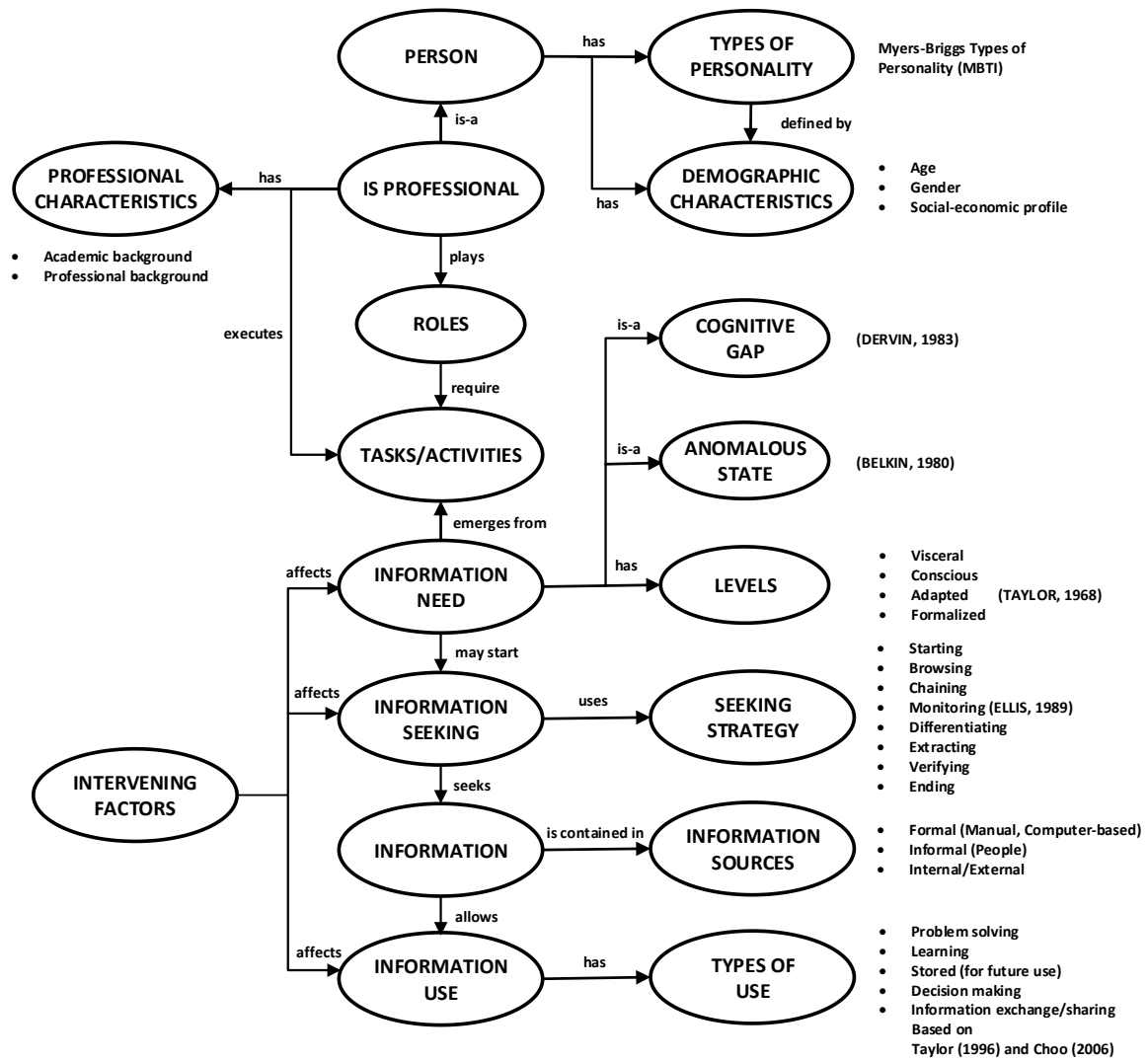


Figure 3. Conceptual Model