

Using Ontological Modeling in a Context-Aware Summarization System to Adapt Text for Mobile Devices

Luís Fernando Fortes Garcia^{1,3}, José Valdeni de Lima², Stanley Loh^{1,4},
and José Palazzo Moreira de Oliveira²

¹ Lutheran University of Brazil, Department of Computer Science, Av. Farroupilha,
8001, Canoas - RS, Brazil - 92425-900

² Federal University of Rio Grande do Sul, Institute of Informatics, Av. Bento Gonçalves,
9500 Porto Alegre – RS, Brazil, 91501-970

³ Faculdade Dom Bosco de Porto Alegre, Department of Computer Science,
Rua Marechal José Inácio da Silva, 355, Porto Alegre – RS, Brazil – 90520-280

⁴ Catholic University of Pelotas, PPGINF, R. Félix da Cunha, 412
Pelotas - RS, Brazil - 96010-000

luis@garcia.pro.br, {valdeni,palazzo}@inf.ufrgs.br,
sloh@terra.com.br

Abstract. This paper presents a context-aware text summarizer based on ontologies intended to be used for adapting information to mobile devices. The system generates summaries from texts according to the profile of the user and the context where he/she is at the moment. Context is determined by spatial and temporal localization. Ontologies are used to allow identifying which parts of the texts are related to the user's profile and to the context. Ontologies are structured as hierarchies of concepts and concepts are represented by keywords with a weight associated.

Keywords: context-aware summarization, ontologies, adaptation, mobile computing.

1 Introduction

Nowadays, people are seeking for casual learning in academic and business organizations. Casual learning differs from formal courses because apprentices look themselves for content and teaching material without having to register or even pay for that. With the growing use of electronic media, apprentices can obtain qualified content in websites, digital libraries and even in CD or DVD materials. However the huge amount of information available electronically all over the world and in different media causes the information overload and thus makes difficult the learning process.

One alternative to minimize the overload is the use of automatic text summaries generated from original documents. This contributes to the reduction of information overload, since only a selection of the most relevant content is exhibited and people can better select information sources where to deepen.

A special case of text summarization is due to mobile computing. Mobile devices such as cellular phones, handhelds, smart phones, pocketPCs, palmtops and some

notebooks have limitations in display size, storage and processing. When people use these devices, they need summarized information that maintains the relevancy and quality of the information as in its original sources but that regards the constraints of the mobile device.

Other problem with traditional text summarization is that results are produced independent of the context of the user. In the majority of the summarization works, the same summary is generated for different people. Summarizers do not consider the different interests that people have. Furthermore, traditional summarizers do not hold the possibility of people using devices in different situations. The same person may play different roles and then he/she may have different interests depending on the moment. The combination of a person and a situation defines a specific scenario.

For these reasons, text summarization should concern questions related to where the user is and what is he/she interested on in order to appropriately generate the summary. This situation is known as “context-aware computing”. Schilit & Theimer (1994) use the term to represent an emerging area to investigate applications that adapt themselves according to the user’s location, to a group of people, to object near the user or to changes occurred with the objects along the time. These conditions compose the context of the application or of the user. According to Chen (2005), “Context is a description of the situation and the environment a device or a user is in”.

Schmidt et al. (1998) describe 6 factors that influence the context: the information about the user, the social environment, the user’s tasks, the location, the infrastructure and the physical conditions. According to Dey (2001), there are 4 kinds of context: the computational context, the user context (related to the social situation of the user), the physical context (related to the environment conditions) and the temporal context (date, time and seasons).

Systems and devices have to adapt to the context, that is, they must alter their behavior (internal tasks and communication interfaces) according to changes occurred in the context (Henricksen et al., 2002). In this sense, summarizers must consider the context to elaborate an extract from a text, generating a different summary for each user and context, according to the target user’s interest and the conditions where he/she is at the moment.

This paper presents a system that generates context-aware text summaries for mobile devices users, modeling the context of the user (interest and situation) with the help of domain ontologies. In this paper, context-aware summarization is defined as the use of additional information, associated with spatial localization and user’s interest, in the method for determining the relevance of extracts to be used for generating a text summary. Domain ontologies are employed to describe and represent the user’s interest and the user’s situation (localization or moment), so that the summarizer is able to generate a different summary according to the combination user + situation. This combination (a user plus a situation) is considered a unique context.

The paper is structured as follows. Section 2 presents related works. Section 3 discusses the theoretical aspects of the modeling and the structure of the domain ontology utilized in the context-aware summarization system. Subsections 3.1 and 3.2 present details of the system (architecture and the summarization process). Section 4 presents concluding remarks.

2 Related Work

The adaptation of texts for mobile devices was discussed in many works. Gomes (2001) proposes heuristics that make possible to display long documents in size-limited devices, with no damage to content understanding. Corston-Oliver (2001) presents an approach to the adaptation of texts that must be displayed in small devices, based on text compaction, generating a telegraphic representation of each sentence by excluding some elements. Buykkokten's proposal (2001 and 2002) for text summarization is implemented in five methods, in which each web page is split in semantic textual units that can be partially displayed. The user can explore successive portions of text in different levels, according to his/her particular needs. The adaptation developed by McKeown (2001), McKeown et al (2003) and Muñoz (2003) takes into account the user's profile, when adapting contents to mobile devices. However, summarization does not consider contextual information, since it is limited to only consider the user's profile (patient or physician) to generate different summaries.

Recently ontologies are being used to structure information. An ontology is a kind of knowledge about the domain and can help in representing the user's profile. For example, Middleton et al. (2003) use a topic ontology to represent interests of users. Topics are extracted from CORA, a domain ontology about Computer Science. Documents are classified in topics by using the k-NN method; the topics associated to the user are those related to documents browsed by users. The work also uses the hierarchy of topics to infer new topics of interest.

Gauch et al. (2003) also use concepts from an ontology to represent profiles. The web pages visited by the user are automatically classified into concepts of an ontology. Profiles are generated from the analysis of the behavior of the user, specifically the content, length and time spent on each visited page. The ontology is based on subject hierarchies from Yahoo, Magellan, Lycos and the Open Directory Project. Only top concepts in the hierarchy were used, leading to more general areas of a user's interest.

The missing point in the cited works is that none of them combines text summarization, profiles, ontologies and context-awareness. The goal of the present paper is to discuss the use of ontologies for representing context (user's profile and situations) in a context-aware text summarizer.

3 Ontologies for Context-Aware Summarization

Users may have different interests or preferences according to conditions where they are. For example, when someone is physically present at the stock exchange, we could consider natural that he/she was provided with summaries about the financial market. However, when he/she goes to a shopping mall, the new context changes the summarization focus to other issues concerning products, sales and movies.

Contextual information is useful to indicate "what" should be delivered to the user, "where" and "when". The user's profile determines which information is interesting for the user. However, depending on where the user is and when (what moment or time), the interest may be biased to a different subset. Thus, information delivered to

user depends on 2 sets of features: information about the user's profile and information about the user's condition at the moment (local and time).

In this work, profiles are persistent interests. We believe that interests represented in the profile remains stable for a short time. Contexts are related to physical or geographical places (spatial localization) and to special conditions of the user at a moment (temporal localization). It is interesting to note that the same context may occur in different places; for example, a context defined as "vacations" may occur in a beach house or in a mountain hotel. Contexts are also determined by temporal information. For example, "beginning of the day" may be a context that may occur at a specified time during the week (6:00 pm) or in an interval in the weekend (between 9 and 10 pm). This context may be used to deliver news for the user. Note that this context may be independent of the physical place ("deliver news to user in that time wherever he/she is") or may be dependent ("only deliver news in that time if user is in the kitchen").

As context changes, the same user may receive different information according to different situations. For example, in an "office" context, user should receive financial news about the market and, while in "vacations", he/she will receive general news.

The system presented in this paper employs information about the user's profile and information about the current context of the user (temporal and/or spatial localization) to determine the kind of text summary that will be generated to him/her. To represent the different kinds of profiles and contexts, this work uses domain ontologies. A domain ontology is a description of "things" that exist or can exist in a domain, Sowa (2000) and contains the vocabulary related to the domain, Guarino (1998). The domain ontology determines which concepts of the real world are being considered and describes how they can be detected. In this paper, domain ontologies are structured as concept hierarchies with each concept being described by keywords. Associated to each word, there is a weight that determines how much the word identifies the concept (profile or context). These keywords will be used to determine which parts of the text will be selected for composing the summary to be delivered to the user. Keywords work like text filters; only phrases that contain one of the keywords are considered relevant to the user (this is an adaptation of the "cue-phrases" concept by Paice (1981) As we explain in the next section, the summarization technique evaluates the relevance of each phrase according to the presence of keywords (profile words and contextual words).

The ontology used for profiles is based on categories about general knowledge, following suggestion of Gauch et al. (2003) and Labrou and Finin (1999) that use categories from *Yahoo!* as an ontology to describe content and features of items. Only high level areas (top topics) are being considered at the moment. Keywords associated to each concept in the ontology and their weights were defined by experts in each area. Table 1 presents part of the ontology employed in the current system for describing profiles, including some terms that identify each profile. The ontology for describing contexts was created manually by the authors. The definition of terms and weights was made by the authors. Table 2 presents part of the ontology employed in the current system for describing contexts (temporal or/and spatial localization). In table 2, some terms employed to identify each situation are shown. These examples do not show the complete ontologies; only part of the concepts is showed and only a few terms were listed for example purpose.

Table 1. Part of the profile ontology

Profile	Keywords
Architecture	house, flat, furniture, decoration, renovate, architecture, curtain
Education	Student, teacher, professor, learning, teaching, school, university
Computers	Computer, software, hardware, cpu, internet, modem, web, ADSL, virus
Fashion	Boot, t-shirt, pants, shoe, coat, tie, fashion, model, clothes, wear
Health	Physician, patient, disease, diagnosis, ambulance, nurse, hospital, clinic, medication, surgery, diet

Table 2. Part of the context ontology

Situation	Keywords
Office, business	deadline, delay, schedule, chief, officer, office, business, technology, dolar, stock, wall street, task
Shopping	buy, shop, delivery, liquidation, discount, sell, sold out, credit card
Vacations, Day off	Vacations, rest, summer, beach, day off, holyday, weekend, entertainment
Home	Television, TV, rest, home, dinner, kitchen, cuisine, talk show

At the moment, the ontologies embedded in the summarizer do not include inference rules. However, as the goal of this paper is also to propose an ontological modeling of context, we will discuss this issue now. One way to extend the ontology is incorporating inference rules over the structure of concepts. The profile ontology may accommodate rules about cross interest. For example, if a user likes “text mining” topic, it is likely that he/she will be interested in issues related to “data mining”.

In the context ontology, we can have rules about the combination of situations. For example, we can define that “summer vacations” is an aggregation of the context “summer” with the context “vacations”. In this case, the aggregated concept (the new context) inherits the terms from both child concepts. The process is similar to the behavior of object-oriented models. In another case, someone can define that context X is incompatible with context Y (for example, “work” and “leisure”). In this case, information about one context may restrict the information about the other context and intersections must be disregarded. In the given example, if the user is in a “work” context, the summarizer must eliminate whatever information about “leisure”.

A special case of inference rule is the combination of the two ontologies (profiles and contexts). Information about one concept in one ontology may be used to infer concepts in the other ontology. For example, if the user has in his/her profile that “he/she is only interested in books”, we can infer that he/she will never be in the “cinema” context. Similarly, we can define a rule such that if the user is habitually in the “cinema” context, then he/she should have “movies” associated to his/her profile.

3.1 Architecture of the System

Basically, the system receives information from a mobile device, including the user identification and the context where the user is at the moment. Then the system retrieves contextual words according to the user’s profile (the user’s profile is stored in a database of profiles) and words associated to the identified context (according to the

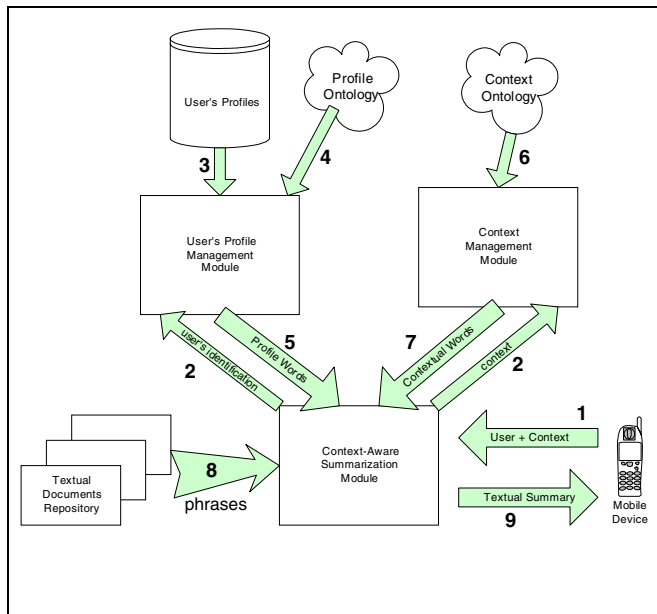


Fig. 1. System architecture

context profile). The final task is to select phrases from textual documents stored in the repository. Each document is processed at one time. The process of filtering phrases is discussed in the next section. For each document, one summary is elaborated and delivered to the user's mobile device. We assume that there is a software module in the mobile device that is responsible for showing the textual summary at the device display.

The system consists of three main modules: (i) User's Profile Management Module; (ii) Context Management Module and (iii) Context-Aware Summarization Module (fig.1). The information flow identified in the figure is described below:

1. Receives information about the user and his/her context
2. Informs User's Profile Management Module and Context Management Module
3. User's Profile Management Module identifies the user's interest in his/her profile
4. Profile Management Module retrieves keywords associated to the corresponding profiles
5. Profile Management Module returns keywords and weights
6. Context Management Module retrieves keywords associated to the corresponding context
7. Context Management Module returns keywords and weights
8. Summarization module separates phrases from the text and ranks the phrases according to the relevance formulae
9. Summarization module sends a summary to the user

The mobile device, where the summarized information will be displayed, is responsible for obtaining information related to spatial localization, temporality and user's identification. For space limit reasons, this paper will not discuss how this information is identified by the mobile device or with the help of other device. We assume that there is a device (hardware or software) that identifies the context (local and time) where the user is in. We can assume that some kind of hardware or software is able to recognize the user by its device and, using a geographical information system, it can determine the place where the user is. Local time may also be informed as contextual information.

The document repository contains the textual documents that will be summarized with the context-aware approach. For each document in the repository, a summary is constructed. Documents could be previously selected and only some of them could be analyzed, but we do not discuss this in this paper.

The database with users' profiles stores information about users and the profiles associated to them (one or more). A profile represents the user's interest in terms of subjects or themes. The user's profile is dynamic and is constantly updated from information about the user. This paper will not discuss how the profiles are created and maintained, but reader can imagine a module that collects explicitly information from the user through questionnaires, forms or interviews or that observes the user's actions in other systems and defines the profile that best match the user's behavior. An extension of this model could include other information as mood, interface preferences and cognitive data.

The profile ontology helps in identifying words related to each profile. These words will be used by the summarization module jointly with context words. The context ontology has definitions about which contexts exist and also the terms related to each context. These context words are used in the summarization process. The user's profile management module receives the user's identification and searches in the profile database for profiles associated to the user. Then, it retrieves from the profile ontology the words associated with the respective weights.

The context manager module is responsible for retrieving contextual words from the context ontology, based on the context identified by the mobile device. The words and their respective weights are then passed to the summarization module.

The context-aware summarization module implements an automatic text summarization technique improved with the inclusion of contextual words supplied by the user's profile management module and by the context management module. The summarization process is detailed in the next section.

The architecture is open, so the summarization module can be implemented and/or expanded with different summarization techniques and with the addition of new kinds of contextual information.

3.2 The Summarization Process

The summarization process starts by receiving contextual words (from the Context Management Module) and profile words (from the User's Profile Management Module) and then picks each document from the Document Repository to generate a summary for each document.

The context-aware summarization approach is based on the idea that the presence of some words in a text demonstrates or are clues that this text is about certain theme or subject. This is the fundament of extractive summarization techniques that select sentences from the original text to compose the summary. The process of selecting sentences evaluates the relevance of each phrase by the relevance of the words that compose it. The weight of each word may be determined by different methods as for example the frequency of the word in the text, the presence in titles and subtitles, etc.

The process proposed in this paper increments the value of sentences extracted from traditional methods of summarization by adding relevance values related to the user's profile and his/her context. As a given word tends to be more significant or frequent in a certain context than in others, its presence in a sentence may increase the relevance of this sentence for users interested in that context. This makes some parts of the text more relevant than others.

It is important to remember that the selection of contextual words and profile words are made previously. Profile words are those related to the profile associated to the user in question, according to the database of profiles and to the profile ontology. Contextual words are those associated to the identified context, according to the context ontology.

The context-aware summarization technique, that calculates the relevance of a sentence, was created based on the TFISF algorithm Larocca Neto (2000). This algorithm was adapted for the present study to incorporate contextual information. The original TFISF algorithm calculates the importance of a word w in a sentence s by the following formulae (1):

$$\boxed{TFISF(w, s) = TF(w, s) \times ISF(w)} \quad (1)$$

where $TF(w, s)$ is the number of times the word w occurs in the sentence s , and $ISF(w)$ is the inverse frequency, calculated with the following formulae (2):

$$\boxed{ISF(w) = \log\left(\frac{tam(s)}{SF(w)}\right)} \quad (2)$$

where $SF(w)$ is the number of sentences in which the word w occurs, and $tam(s)$ is the total number of words that compose the sentence s .

The final calculus of the relevance value of the sentence s with the addition of context-awareness is shown in the following formulae (3):

$$\boxed{TFISFca(s) = \sum_{j=1}^n \left[TFISF(w_j, s) \times IP(w_j) \times IC(w_j) \right]} \quad (3)$$

where $TFISFca(s)$ is the relevance value for the sentence s according to the context-aware technique presented in this paper; and $TFISF(w_j, s)$ is the relevance of the word j in the sentence s according to the original TFISF algorithm; and $IP(w_j)$ is the weight of the word j given by the profile ontology; and $IC(w_j)$ is the weight of the word j given by the context ontology; being n the number of significant words in the sentence s .

If the word j is not present in the profile of the user, the value for $IP(w_j)$ is 1. If the word j is not present in the current context, the value for $IC(w_j)$ is 1. The weights of

the words in the ontologies range from 2 to 10. For the moment, we are assuming one unique profile as valid at a moment.

In this technique, besides the traditional TFISF calculus for relevance of sentences, a sentence that contains profile or contextual words have its relevance value multiplied by the weights associated to these words in the ontologies, increasing the chances of that sentence to participate in the final summary.

If one word in a sentence is not in the user's profile or in the current context, its original relevance value is maintained and its original contribution for the sentence selection is preserved.

The process excludes words known as stopwords, such as prepositions and articles. Thus, only significant words are considered in the relevance calculus.

The result of the process (the textual summary) is composed by the most relevant sentences, those with the highest relevance values, after the calculus is made for all sentences in the textual document.

4 Conclusions

Experiments carried out with the system in a real situation. Texts were gathered from a Brazilian periodic (*Veja*). Ontologies were manually elaborated. The profile ontology has 10 profiles (economy, fashion, sports, politics, medicine, religion, tourism, police, arts, education) while the context ontology was created with 5 contexts (home, shopping, vacations, work, supermarket). Approximately 30 words were associated to each concept in the ontologies. Summaries were generated by the proposed method for each combination of one profile and one context. Human judges rated each summary according to the combination, analyzing relevance and coherence. The results indicate that 67% of the summaries were rated as relevant to the combination of profile+context and 80% of summaries were judged as coherent to the original text by maintaining information coherent.

Although the promising results, we must carry out another experiment with an objective evaluation. The subjective evaluation is important but suitable to mistakes. The problem is to find a measure appropriated for the purpose (personalized and context-aware summarization of texts).

During the work, we found some obstacles. The elaboration of the ontologies by hand is time-consuming and error-prone. The definition of concepts for each ontology is theme of another paper. We are investigating the use of machine learning methods to semi-automatically identify concepts determine. Another problem is the definition of words and their weights for each concept. We intend to use text mining methods such as TFISF, Latent Semantic Indexing and Bayesian networks to automatically find words and weights.

Other difficulties related to the ontologies include the ambiguity problem (when a keyword appears in more than one concept, generating summaries out of the context) and the lack of a stemming method for preprocessing of the words.

A special case for future works is the study of the coherence among the phrases that compose a summary. In the current stage, summaries are generated by the concatenation of the selected phrases. This can generate coherence problems when passing from one phrase to the next.

In future works, context must be determined by automatic mechanisms and profiles will be learned by intelligent software tools.

The main contributions of this work are the development of an open architecture for context-aware systems and the automatic text summarization process based on context and personalized information. Results from the experiment lead us to the conclusion that the proposed method is promising and can provide more focused information.

The application of the proposed method in learning environments and situations is well appropriated since apprentices can receive personalized and summarized information wherever they are. This can reduce the time for knowledge acquisition and minimize the overload problem. Summaries are oriented to interesting themes and momentary situations, freeing humans from the need for selecting text parts and from giving information about their interests.

References

1. Anderson, C., Domingos, P., Weld, D.: Web Site Personalizers for Mobile Devices. In: International Joint Conference on Artificial Intelligence, Proceedings... [S.l.:s.n.], vol. 17 (2001)
2. Buykkokten, O., et al.: Efficient Web Browsing on Handheld Devices Using Page and Form Summarization. *ACM Transactions on Information Systems*, New York 20(1) (January 2002)
3. Chen, A.: Context-Aware Collaborative Filtering Systems: Predicting the User's Preferences in Ubiquitous Computing. In: CHI 2005. Proceedings Conference on Human Factors in Computing Systems, pp. 1110–1111 (2005)
4. Corston-Oliver, S.: Text Compaction for Display on Very Small Screens. In: Meeting of the North American Chapter of the Association for Computational Linguistics, Proceedings... [S.l.:s.n.], vol. 2 (2001)
5. Dey, A.: Understanding and Using Context. *Personal and Ubiquitous Computing*, London 5(1), 4–7 (2001)
6. Gauch, S., Chaffee, J., Pretschner, A.: Ontology-based personalized search and browsing. *Web Intelligence and Agent System* 1(3-4), 219–234 (2003)
7. Gomes, P., et al.: Web Clipping: Compression Heuristics for Displaying Text on a PDA. In: MOBILE HCI, 2001. Proceedings.. [S.l.:s.n.] (2001)
8. Guarino, N.: Formal Ontology and Information Systems. In: FOIS 1998. International Conference on Formal Ontologies in Information Systems, Trento, Italy, pp. 3–15 (1998)
9. Henriksen, K., Indulska, J., Rakotonorainy, A.: Modelling Context Information in Pervasive Computing Systems. In: Mattern, F., Naghshineh, M. (eds.) *Pervasive Computing*, LNCS, vol. 2414, Springer, Heidelberg (2002)
10. Labrou, Y., Finin, T.: Yahoo! as an Ontology - using Yahoo! categories to describe documents. In: CIKM 1999. 8th International Conference on Knowledge and Information Management, Kansas City, MO, pp. 180–187 (October 1999)
11. Larocca Neto, J., Santos, A.D., Kaestner, C.A.A., Freitas, A.A.: Document clustering and text summarization. In: Proc. 4th Int. Conf. Practical Applications of Knowledge Discovery and Data Mining, pp. 41–55 (2000)
12. McKeown, K.: PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information. In: JCDL. Proceedings.. [S.l.:s.n.] (2001)

13. McKeown, K., Elhadad, N., Hatzivassiloglou, V.: Leveraging a Common Representation for Personalized Search and Summarization in a Medical Digital Library. In: Joint Conference on Digital Libraries, 2003. Proceedings.. [S.l.:s.n.] (2003)
14. Middleton, S.E., Shadbolt, N.R., Roure, D.C.D.: Capturing interest through inference and visualization: ontological user profiling in recommender systems. In: KCAP 2003. International Conference on Knowledge Capture, pp. 62–69. ACM Press, New York (2003)
15. Muñoz, M., et al.: Context-Aware Mobile Communication in Hospitals. *IEEE Computer*, [S.l.] 36(9) (September 2003)
16. Paice, C.: The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. In: Oddy, R., et al. (eds.) *Information Retrieval Research*, London: [s.n.] (1981)
17. Pardo, T., Rino, L.: TeMário: Um Corpus para Sumarização Automática de Textos. São Carlos: Universidade de São Carlos, Relatório Técnico (2003)
18. Schilit, B., Theimer, M.: Disseminating Active Map Information to Mobile Hosts. *IEEE Network*, [S.l.] 8(5), 22–32 (1994)
19. Schmidt, A., Beigl, M., Gellersen, H.: There is more to Context than Location. In: International Workshop on Interactive Applications of Mobile Computing, 1998. Proceedings.. [S.l.:s.n.] (1998)
20. Sowa, J.F.: *Knowledge representation: logical, philosophical, and computational foundations*. Brooks/Cole Publishing Co., Pacific Grove, CA (2000)