

Task-Centered Selection of Learning Material

Alexander Streicher, Natalie Dambier, Wolfgang Roller

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
Karlsruhe, Germany
firstname.lastname@iosb.fraunhofer.de

Abstract: Learning needs in complex working environments call for e-learning systems which intelligently support the learner. As everyday tasks keep getting more and more complex employees consistently have to update their knowledge and adapt to new processes. As a consequence a lot of time has to be spent on research for appropriate help and learning material. The aim is to decrease the time the user has to spend on his hunt for information and to offer him the needed help and learning material in an on-demand manner. We present a new approach for semantic retrieval of learning units taking the working context into account. Basis is an ontology with attached binding weights. A context-aware ranking of help and learning material is generated with a semantic spreading activation algorithm. The gained semantic search results match to the learner's actual situation better than e.g. a pure full-text search, because the underlying ontology-based retrieval is aware of relations in the search domain and uses this knowledge in a way aligned to the learning process as well as to the specific domain. The results are shown in a prototype implementation of an assistance and learning system for Synthetic Aperture Radar (SAR) image interpretation. This work is based on [15] and is extended by new aspects to the retrieval method and a comparison with a full-text search engine.

Keywords: semantic retrieval, e-learning, image interpretation, ontology, semantic spreading activation

I. Introduction

To be able to adapt to new scenarios and challenges as well to solve new problems we must continuously update our knowledge. Lifelong learning enables us to incorporate new information with our existing knowledge base and our experience. Especially nowadays the rapid development of our technology and information aspiring society changes the way we interact with complex systems in our everyday life. Although we apparently have to move along with the changes there are to come most of the time it is a tedious task. We cannot succeed in knowing everything, especially everything which is new. Because of that people are not trained for their tasks only in school or only with specific qualifying training but actually at their workplaces. This is especially true for people who work in complex working environments. Experience alone cannot help them to solve new problems. They have to update their knowledge according to the tasks at hand and incorporate that new knowledge into their own

experience. Complex systems and complicated workflows demand flexible employees who are willing to update their knowledge. E-learning technologies can assist these employees. Assistance and e-learning systems help the user with their on-the-job training. In this paper we present a new approach how to offer the learner the information which is most relevant to his current working context.

The work of an image interpreter, especially in the domain of aerial image interpretation, perfectly fits the description of a complex working environment. The image interpreter must recognize objects (such as vehicles, buildings, site infrastructures etc.) and interpret their meaning based on aerial images. Different sensor and imaging parameters, a high variety in appearance of objects around the globe and time pressure create a challenging working environment. One of the most demanding tasks is the analysis of complex facilities (such as airfields, harbors and industrial installations) based on Synthetic Aperture Radar (SAR) images. SAR is used in a wide variety of application scenarios, e.g. pollution detection, cartography, ice layer and biomass monitoring as well as reconnaissance and surveillance. For the latter the radar images have some advantages over optical images, for instance no need for illumination (e.g. from the sun) and it is rather unaffected by weather effects like clouding.

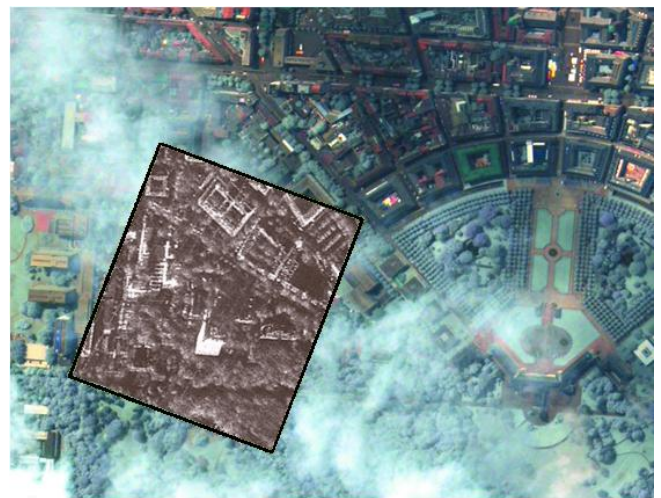


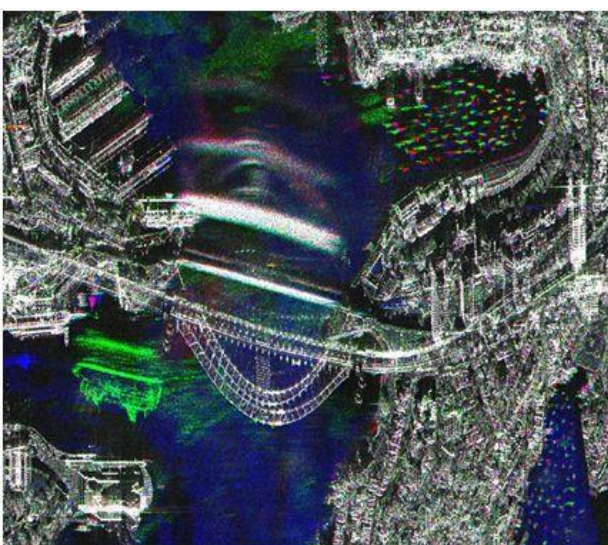
Figure 1. Radar vs. optical image: penetration of clouding (copyright by Cassidian, radar, and Eurimage, optical)

As an example have a look at *Figure 1*. You can see an optical satellite image of downtown Karlsruhe, a city in the South of Germany. At some parts the area is clouded and the optical image is obviously affected by that. However, a radar sensor is able to penetrate the clouding and present the viewer with the area beneath the occluded area. Obviously this is of high interest for reconnaissance.

However, SAR has also some disadvantages. The radar signatures of objects in the resulting image differ significantly from those in an optical image. For instance see *Figure 2*. You can see the optical image of the famous harbour bridge and its surrounding area in Sydney, Australia. In the radar signature there are a lot of particular effects which make the image hard to interpret. The objects like the bridge or the buildings are flipped downwards towards the origin of the radar beam (layover effect). The wide white bar on top of the bridge arc is actually the bridge's bottom. This is because the radar beam is being reflected by the water (layback effect).



(a) Optical image (copyright by Google Inc.)



(b) Radar signature (copyright by Infoterra GmbH)

Figure 2. Optical image and radar signature of the harbor area of Sydney, Australia

The interpretation of these effects is crucial for the image interpretation task. To ease the identification and classification process assistance systems are being developed which help the user to classify objects in an image. Although these systems can offer a wide variety of tools, e.g. for image processing, image annotations or automatic classification [3] [8], the human-factor in the interpretation process still remains the essential element of correct and sound interpretation.

As the user constantly interacts with the assistance system during the interpretation process, the system is well aware of the current state of the task. For example, it collects information on objects which have already been recognized. This knowledge about the current state of the task can be used in order to provide the user with useful help and learning material, e.g. learning units of an e-learning system. This can be applied as soon as the user reaches the limits of his current knowledge and experience. We present a new approach for semantic retrieval of learning units depending on the current task context. Our approach is based on an ontology with attached binding weights and semantic spreading activation [9]. It provides the user with qualified learning material which is intelligently retrieved based on the current working situation. This is in contrast to previous systems where the retrieval of information is solely based on text retrieval methods, thus considering a limited search space only.

The preliminary results of our work are shown in a prototype implementation of an assistance and learning system for SAR image interpretation. The aim is to optimally assist the image interpreter in his work by offering appropriate learning units for search objects in an image.

II. Related Work

In this paper an ontology-based spreading activation retrieval algorithm is presented which enables the e-learning system to find learning material tailored to the user model.

Using ontologies in e-learning systems and linking assistance systems is a growing field of research [1] [11]. Ontologies can be used to model teaching knowledge [5] as well as to exploit Semantic Web techniques to enable for instance machine-based logical reasoning [13] and semantic search [12]. However, in the context of true information retrieval pure semantic search lacks in the ability to rank the results which renders the search process as plain data retrieval only. Combined with weights in the semantic network a mechanism based on the spreading activation principle [7] is able to produce scores for each accounted concept to enable a ranking of the results. Spreading activation in information retrieval can be seen in [6] and [9]. An approach similar to ours but without reference to context-aware e-learning is shown in [17] in which a search architecture is presented that combines classical search techniques with spreading activation techniques to execute semantic searches in websites.

Rather than using keywords as user input for the semantic search it is possible to provide the search terms automatically by a preceding system, e.g. by an assistance system. This can be seen as intelligently interlinking multiple assistance systems. Interlinking of assistance systems and e-learning

systems has been presented by [11] and [15]. They give a description of how to interlink an assistance system [4] for use in image interpretation of SAR images with an e-learning application.

III. Semantic Data Model (Ontology)

The search space for the presented semantic retrieval is spanned using an ontology which is based on the principle of the Simple Knowledge Organization System (SKOS) [16]. SKOS provides an elementary vocabulary to describe basic structures of concept schemes. This structure is easy to handle and, if consequently used, it leads to interoperability between different systems. The SKOS representation is transformed to F-Logic which allows specification of complex rules and efficient reasoning [10].

As an example for reasoning see Figure 3. Initially the concepts “Geometric Shape” and “Storage Tank” are not defined as to be related to each other. The relations between “Building” and “Storage Tank” and “Geometric Shape” are defined to be transitive. Because of that it can be inferred that “Geometric Shape” is transitively related to “Storage Tank”. The actual type of relationship is then defined by the underlying logic in the ontology. A similar reasoning can be done for the learning unit “Petrol, Oil and Lubricants” which can be inferred as to be related to “Geometric Shape”.

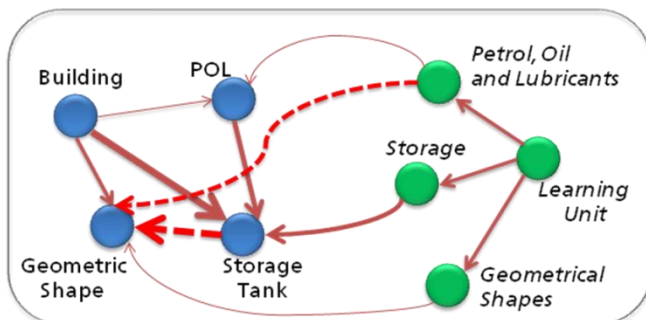


Figure 3. Simple example for reasoning in a semantic network. Solid lines are defined relations, dashed lines indicate relations automatically inferred by a reasoner engine

Basis for the search process is a domain ontology which describes the topic of the learning units. In the example case of SAR image interpretation it is an ontology of airfields enriched with simple geometrical aspects. The domain ontology consists only of the one concept *skos:Concept*; all other items are instances of this concept. The relations between instances are defined by the SKOS-relations *broader*, *narrower* (inverse to *broader*) and *related* extended with the self-defined relation *hasPart* to construct a partonomy. Furthermore an identification label as well as synonyms and translations are introduced to offer a broader search space and to provide internationalization. The second part of the ontology model describes the learning units. Each learning unit is an instance of the concept *Document*. The main annotation relations are *hasPrimarySubject*, *hasSubjectTags* and *hasKeyword(integer)*. The connection to exactly one concept of the domain ontology is established by the relation *hasPrimarySubject*. The relation *hasSubjectTags*

assigns further topics and may link to several concepts – the number is not limited. But in practice it is not reasonable to assign more than ten tags, because a learning unit should be designed to explain the topic precisely and to focus on one topic only. These tags should be items mentioned in the learning unit supporting the main subject. Figure 4 sketches an example ontology based on the described schema including domain and content ontology.

The relation *hasKeyword(integer)* is not explicitly defined a priori, it is inferred during the reasoning process. The assigned integer value is the calculated binding weight for any concept accessible through relations starting from the learning unit by spreading activation. There may be further relations defined by a didactic ontology (e.g. as in [18]) describing the didactic dependencies to other units as well as properties like media type. These relations may be used for the visualization of search results but are not essential for retrieval.

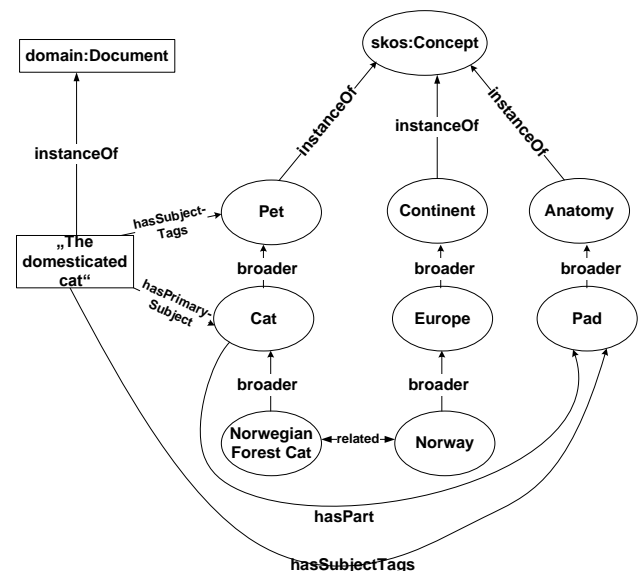


Figure 4. Example of the learning unit “The domesticated cat” as instance of the content ontology linked to concepts of the domain ontology

IV. Semantic Retrieval

The primary objective of the semantic retrieval in the current context of retrieving learning units is to intelligently find those learning units which match the user’s needs best. To find semantically relevant concepts the search process makes use of the ontologies’ semantic net structure and applies the technique of spreading activation.

A. Semantic Spreading Activation

The spreading activation mechanism originates in cognitive psychology [7] to model spontaneous associations when the brain recognizes a word and activates other concepts linked to that term. In information retrieval spreading activation can be applied to expand the search space [2] [9] or to retrieve linked concepts [6].

In this work the spreading activation process is applied to a semantic net, and there to labeled nodes (concepts) and

weighted edges (relations). When activated, the weight, or “activation energy”, of each activated node is propagated through the network to their linked nodes. In our semantic net we use real valued weights $w \in \mathfrak{R}$, where $0 \leq w \leq 1.0$. The weights can be discounted by multiplication as the activation spreads through the network rendering the neighboring nodes most important and the most distant ones as irrelevant. Intuitively this makes sense since the most relevant concepts are most likely seen in close neighborhood and the most irrelevant ones are located at the other end.

In a recursive fashion the propagation of the binding weights is computed for a node n_i and the linked node n_{i+1} as

$$O(n_{i+1}) = O(n_i) \cdot w[r(n_i, n_{i+1})]$$

$O(n_{i+1})$ denotes the output $O: N \rightarrow \mathfrak{R}$ of the linked node n_{i+1} and $O(n_i)$ the output of the preceding node n_i . Let N be the set of all nodes in the network and R the set of all relations between the nodes. An edge between two connected nodes $n_i, n_{i+1} \in N, i \geq 0$ is defined as a relation $r: N \times N \rightarrow R$. The function $w: R \rightarrow \mathfrak{R}$ yields the binding weight for a single relation $r \in \mathfrak{R}$. The base case for the starting node n_0 is defined as $O(n_0) = 1$. An illustration is given in Figure 5 for a simple network of just two nodes.

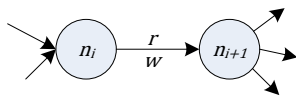


Figure 5. Two nodes in a network connected by relation r with associated weight w

Various strategies have been proposed when to stop the propagation process [17], e.g. stop when a specific concept is hit (concept type constraint) or when the output’s lower limit is hit (distance constraint). Latter is used in this work. The spreading stops when the node output drops below a given threshold T , i.e. $I < T$.

Here, an ontology is used as the semantic network. The concepts are the nodes of the network whereas the properties or relations are the edges. Semantic spreading activation takes into account the meaning of the relations. Thus, in combination with a reasoner, the weights are semantically supplemented by means of their logical correlation. Inference in the ontology can yield to new relations between the nodes and therefore enhancing the search space drastically. As an example see Figure 6 where in addition to the explicitly stated relations (the solid lines) a new relation between LU_1 and $C3$ has been inferred (the dashed line) using an ontology reasoner.

B. Semantic Retrieval

For each relation in the SKOS-based ontology weights are introduced which influence the rank of the found documents in the retrieval result (Table 1). These weights were initially chosen due to the following considerations: the primary subject is what the learning content is about and usually part of the document’s heading, thus *hasPrimarySubject* gets the

highest binding weight. The subject tags (relation *hasSubjectTags*) are equally directly associated to the unit and should therefore be considered more relevant than any other concept reached by the spreading activation process. Hence the relations between the domain ontology concepts are weighed less than the content ontology relations.

Regarding the domain ontology the relations *broader* and *hasPart* are the ones with the highest weight. Both of them have a close correlation to the origin term. For a specific relevant concept it is often helpful to take the more general concept (relation *broader*) into consideration, too, to get a more complete overview. And, for *hasPart*, if something as a whole is in focus, the parts of it may help to understand it better. Whereas *narrower* - although similar to *hasPart* and inverse to *broader* - may lead to a more specific term that is less helpful to solve the learner’s actual problem. For example if the topic of interest is a cat it may be interesting that a cat has something to do with pets (*broader*) as well as the fact that a cat typically has four pads, a tail and long whiskers (*hasPart*). But one cannot automatically assume that the learner needs information exactly about the Norwegian Forrest Cat (*narrower*). The relation *related* can be seen as in between. Relation is not such a strong binding as a paronymy but may be much more helpful than the more specific concept. The initial weights were tested on an excerpt of a learning course and led to the expected results, so that only little modifications were necessary. Table 1 shows the experimentally determined weights.

Relation r	Origin	Weight w
domain:hasPrimary-Subject	content ontology	1.0
domain:hasSubjectTags	content ontology	0.9
domain:hasPart	domain ontology	0.8
skos:broader	domain ontology	0.8
skos:narrower	domain ontology	0.7
skos:related	domain ontology	0.75

Table 1. Weights for the semantic spreading activation process

The primary objective of the presented retrieval is to find relevant learning units and to rank the results based on their relevance regarding the current working situation. The algorithm calculates a binding weight for attached concepts for each learning unit depending on how the concepts are related to each other using the weights defined in Table 1. According to the spreading activation principle the terms in the ontology are activated and its activation energy is passed through the network degenerating in accordance with the weights of the relations. The result is a list of learning documents (learning units) where the most semantically relevant documents are ranked first.

Knowing the learning context it is even possible to use more than one search keyword. $C = \{c_1, \dots, c_m\}$ is the set of concepts of the domain ontology, $S = \{s_1, \dots, s_n\}$ is the set of all search keywords and a subset of C ($S \subseteq C$). $LU = \{LU_1, \dots, LU_m\}$ are the learning units and R is the set

of relations. The set of paths $Paths_{LU_j, c_i} = \{P_1, \dots, P_l\}$ are all possible connections from learning unit LU_j to concept c_i regarding the associated relations. The binding weight $B(LU_j, S)$ of a learning unit is calculated as

$$B(LU_j, S) = \sum_{i=1}^n b_{LU_j}(s_i)$$

with $s_i \in S$, and $b_{LU_j}(s_i)$ is the factor of each conducted spreading activation regarding every search keyword.

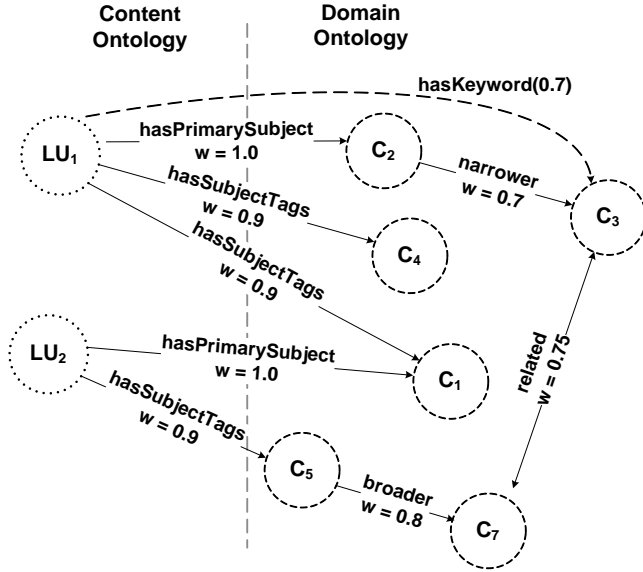


Figure 6. Learning Units in relation to the domain ontology concepts with weighted relations. These weights are propagated through the network

The particular binding weights of each search term are summed up (not multiplied) to guarantee variety in search results. Multiplication would result in excluding learning units where one or more search terms are not linked in any way to that learning unit. To prevent influence of concepts with small binding values on the ranking only concepts with a binding value equal to or greater than 0.4 are considered:

$$b_{LU_j}(s_i) = \begin{cases} \alpha_{LU_j}(s_i), & \alpha_{LU_j}(s_i) \geq 0.4 \\ 0, & \text{else} \end{cases}$$

Thus $\alpha_{LU_j}(s_i)$ expresses the binding weight of the search keyword s_i according to the learning unit LU_j . It can be calculated evaluating the paths $Paths_{LU_j, s_i}$ using spreading activation, i.e. multiplying the individual relation weights $w(r)$ on the paths. If there are several different paths from the learning unit LU_j to the search term s_i , the maximum value is used, i.e. use the path with the highest “energy”:

$$\alpha_{LU_j}(s_i) = \max_{P \in Paths_{LU_j, s_i}} \left(\prod_{r \in P} w(r) \right)$$

Figure 6 gives an example for concepts related to a learning unit. For the concepts C_3, C_4, C_7 the binding weights are computed as follows:

$$\begin{aligned} b_{LU_1}(C_3) &= 1.0 \cdot 0.7 = 0.7 \\ b_{LU_2}(C_3) &= 0.9 \cdot 0.8 \cdot 0.75 = 0.54 \\ b_{LU_1}(C_7) &= 1.0 \cdot 0.7 \cdot 0.75 = 0.525 \\ b_{LU_2}(C_7) &= 0.9 \cdot 0.8 = 0.72 \end{aligned}$$

For the given example in Figure 6 and an assumed search query $S = \{C_3, C_4, C_7\}$ the binding weight result for LU_1 is computed as $B(LU_1, S) = 0.7 + 0.9 + 0.525 = 2.125$.

Because of semantic retrieval the system is able to find learning units that are relevant. The domain ontology uses dependencies and relations to describe the domain leading to more specific search results compared to simple full-text search. Knowing the working context it may be useful to additionally weigh the binding of search terms according to their relevance for the search inquiry.

V. Application

The realization of the described retrieval process is shown in a work in progress with focus on SAR image interpretation (see also section I). SAR is an imaging technology based on reflections of microwave pulses emitted by a radar sensor. Due to the complex imaging geometry and the very different reflection properties of objects in the microwave band, special training and substantial experience are required in order to be able to identify objects in this kind of images.

A. Involved Systems

In order to improve the training of image interpreters, an e-learning system has been customized for this task [19]. The e-learning system includes a comprehensive course about image interpretation in the domain of radar images for reconnaissance. It covers the very basics starting with physical principles of radar waves and continues to specific content like the actual radar signatures of buildings and vehicles. Figure 7 shows a page of this e-learning system. The page contains some explanatory content explaining the basic principles of basic geometrical shapes, a test item and the possibility of starting a simple radar signature simulator. For various objects (in this case a simple rectangular cube on the left in Figure 7) the sensor angle can be adjusted and the resulting radar signature with the corresponding shadow is then been rendered (as seen in black and white on the right).

To support the image interpreter during the interpretation process, assistance systems for image interpretation are developed. As reliable algorithms for automated object recognition in aerial images are hardly available, these systems are often based on interactive approaches [4]. An example of such a system for the task of image interpretation for infrastructure is shown in Figure 8.

The assistance system supports image interpreters to perform a full analysis of a complex object arrangement, for example it helps to decide whether a radar image shows a civilian or military airfield. Singular objects (buildings, roads etc.) are marked by the user in the image and the system

makes use of a probabilistic scene model to classify the function of the singular objects as well as the type of the overall facility. The classification results are presented to the user as recommendations.

The e-learning application has been developed for education and information transfer in military image interpretation. It provides courses and training content as well as background information for SAR image interpretation.

Both systems go hand in hand in the education and knowledge transfer of image interpreters, and linking these two systems is an obvious objective.

Semantic retrieval enables the image interpretation system to provide the learner with context-adapted learning material. Our implemented prototype interlinks an assistance system with an e-learning application. It combines information provided by the assistance system with existing learning material.

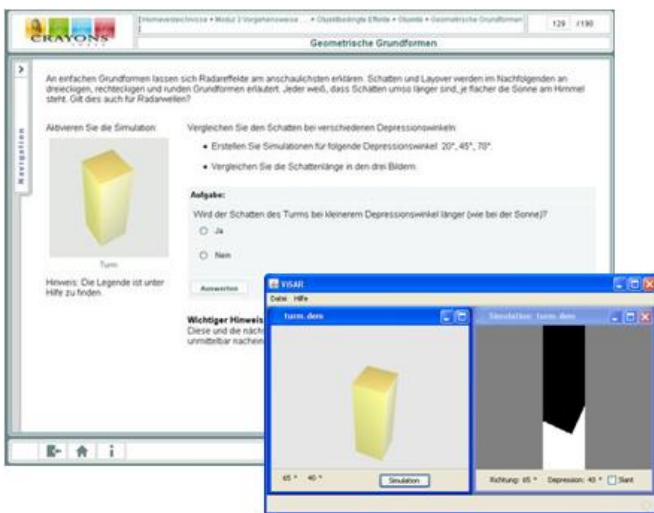


Figure 7. Example page of the involved e-learning system (top left). Displayed is some explanatory content combined with a simple test item. The second image (lower right) shows a training simulator for radar signatures

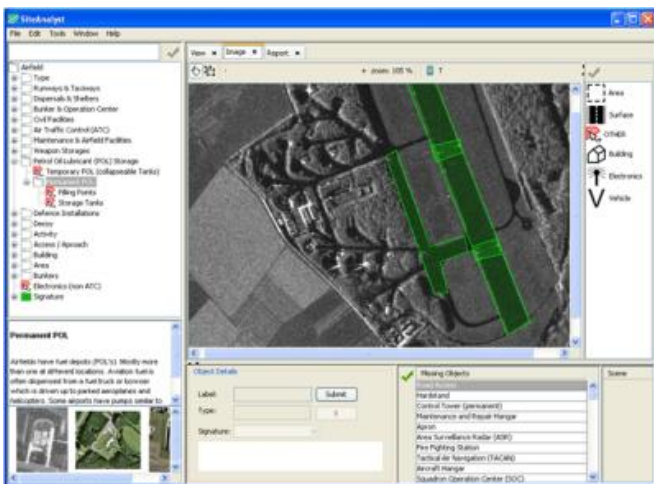


Figure 8. Assistance system for radar image interpretation. The image in the center of the application shows a radar image of an airport and marked areas of interest

The data structures of the described systems were used to populate the domain ontology. The instances were amongst others collected from several existing data sources, e.g. from the taxonomic data structure of the involved assistance system and from already existing ontologies.

B. Intelligent Interlinking with Semantic Retrieval

If the user, i.e. the image interpreter, is at a loss with his knowledge in a current interpretation task, the system should be able to provide this learner with context-adapted on-demand learning material. For this the semantic retrieval algorithm finds the most relevant learning and help material.

The process of how the user interacts with the involved systems described before is shown in the sequence diagram in Figure 9.

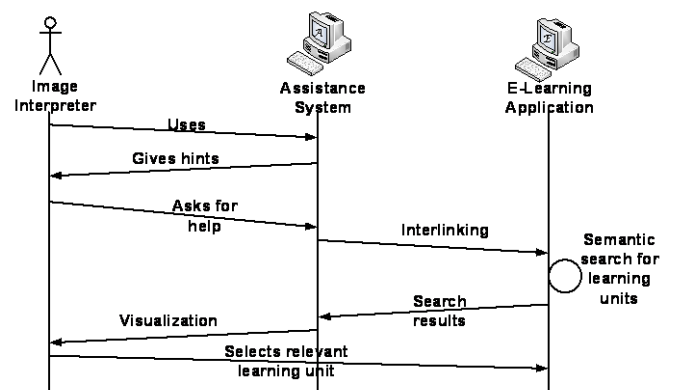


Figure 9. Interaction process. The image interpreter works with the assistance system and asks for help. The assistance system transfers the collected data to the learning system where the context-aware search is performed

When the user needs help for a specific concept the associated term and the probability distributions for all other linked terms are submitted by the assistance system to the e-learning application. The assistance system transfers the so far collected data: already found objects, missing probable objects and of course the actual search object. The e-learning application performs a search on the available learning units which are annotated in the ontology. For each learning unit a binding weight for the received keywords can be calculated. The weights of the relevant keywords are multiplied by a factor related to each keyword source: the actual search term is rated as very important, whereas objects which have already been found get a low factor.

The result of the semantic retrieval is a list of learning units sorted by their relevance according to the spreading activation algorithm as explained before. The concept which has been selected by the user in the assistance system is assigned a higher initial weight to boost it to the top of the result list. However, based on the other concepts and their cumulated binding weights this boosted entry can actually be topped by concepts which are semantically more relevant in the current working situation. More semantic relevance means that they share higher weighted relations to a lot of other concepts which in turn results in higher binding weights.

C. Scenario & Use Case

As an example scenario the image interpreter has to analyze a site infrastructure of an airport. So far he has identified the runway, some barriers and some taxiways, but he is unsure about the other infrastructure components. The identification of some storage tanks for petrol, oil and lubricants (POL) would help him a lot to continue his task. This building is crucial for the further interpretation process because of its distinct characteristics to distinguish the airport type – if it is a civil or military airport. For a civil airport the storage tanks are built above ground in contrast to the military case where they are hidden under ground covered with soil. Figure 10 gives an impression as how these storage tanks appear from an aerial image.



Figure 10. Storage tanks for petrol, oil and lubricants (POL)

As a next step the user selects this item from a list of possible other objects in the scene and asks for help. He expects relevant help and learning material which fits best to his current working context.

D. Result List

To give the user direct access to the help and learning material in the presented e-learning system a sorted list of direct hyperlinks is presented. See Figure 11 for an example output for the current scenario. The entry “Permanent POL” is at the top of the list because on the one hand it is the primary search keyword and on the other hand it has a high semantic binding weight for the current context. The following entries represent learning material for different kinds of airfields where storage tanks are needed as well.

Particularly interesting is the learning unit for “Geometrical Shapes”. This is listed because of a relation between the concept “storage tank” and “cylinder”. As one can easily see in Figure 10 the storage tanks have a cylindrical shape in an aerial image.

E. Comparison to Simple Full-Text-Search

To assess the results of the semantic retrieval approach it has been compared to the results of a full-text search engine. To guarantee objectiveness only the retrieval part has been exchanged. Everything else from the search query term given by the assistance system to the presentation of the results is exactly the same.

For the full-text search engine the widely known *Apache Lucene*¹ search engine was used. This engine allows for

weighting of the query’s input terms. Because the semantic retrieval approach utilizes a similar technique, this feature was used as well for Apache Lucene.

Because full-text search engines work only on the syntactic level the results include only those entries for which there are text-hits in the underlying documents. For the given scenario an example output is shown in Figure 12.

Suggested Learning Units

<input checked="" type="checkbox"/>	Permanent POL (Relevance: 21.7)
<input checked="" type="checkbox"/>	Helicopter Airfield (Relevance: 17.2)
<input checked="" type="checkbox"/>	Transport Airfield (Relevance: 17.0)
<input checked="" type="checkbox"/>	Civil Airfield (Relevance: 15.7)
<input checked="" type="checkbox"/>	Mixed Airfield (Relevance: 15.6)
<input checked="" type="checkbox"/>	Geometrical Shapes (Relevance: 13.0)
<input checked="" type="checkbox"/>	Maintenance/Airfield Facilities (Relevance: 12.5)
<input checked="" type="checkbox"/>	Exercise 2 Access/Approach (Relevance: 10.7)
<input checked="" type="checkbox"/>	Access/Approach (Relevance: 10.3)
<input checked="" type="checkbox"/>	Runways & Taxiways (Relevance: 10.0)

Figure 11. Semantic retrieval results. Listing with relevance bars (on the left) of learning units which are relevant to the user’s current working context

<input checked="" type="checkbox"/>	Permanent POL (Relevance: 1.0)
<input checked="" type="checkbox"/>	Airfield Type (Relevance: 0.9)
<input checked="" type="checkbox"/>	Dispersal Area (Relevance: 0.3)
<input checked="" type="checkbox"/>	Airfield (Relevance: 0.3)
<input checked="" type="checkbox"/>	Control Tower (Relevance: 0.3)
<input checked="" type="checkbox"/>	Wetter- und Navigationsradar (Relevance: 0.2)
<input checked="" type="checkbox"/>	Runways & Taxiways (Relevance: 0.2)
<input checked="" type="checkbox"/>	Image Overlay Check (Relevance: 0.1)

Figure 12. Full-text search retrieval results

For the given scenario the entry “Permanent POL” gets a high score, because it has been found multiply times in the underlying document for the learning unit in the e-learning system. This is as expected. But in comparison to the semantic retrieval approach (see Figure 11) no other semantic relevant content is found. Additionally the relevance of the other entries is much lower compared to the semantic retrieval result. The entries “Geometrical Shapes” and “Maintenance/Airfield Facilities” couldn’t be found because they don’t contain the term “Permanent POL” neither do they contain parts of it or its word-stems. This shows the advantage of the semantic retrieval approach. Related information is found as well and, most importantly, can be ranked by their semantic relevance.

¹ Apache Lucene search engine: <http://lucene.apache.org>

VI. Conclusion and Outlook

In this paper we presented a new approach for semantic retrieval using the user's current working context. The semantic search is executed with a semantic spreading activation algorithm. The described retrieval method leads to context-aware support for a learner in his work process. The semantic search results fit better to his actual situation than e.g. a full-text search, because the underlying ontology-based retrieval is aware of relations in the search domain and uses this knowledge in a way aligned to the learning process as well as to the specific domain. The application is not limited to a specific domain - once implemented it can be used in any context if an adequate domain ontology exists. Moreover the SKOS approach allows for interoperability to other domains, direct reusability and feasible maintenance.

The prototype is still under development. Further research has to be done as how to (automatically) determine the definite weights for the relations and the different kind of keywords. Furthermore the retrieval will be tested on complex data to gain more information on the computability of the algorithm considering different reasoners. The challenge in modeling the domain ontology is not simply to copy the content of a learning unit but to provide an overall view on the domain. Ideally the domain ontology already exists before learning material is developed. This may help the e-learning author to structure his work and allows him to match learning units with the domain ontology at once. Further investigation will be done with text-mining algorithms to allow for semi-automatic text-annotation to help building the ontology. An open field of research is also how to optimally present the learning unit suggestions to the user and how to make the semantic relationships transparent. This will be done using network visualization techniques.

References

- [1] Amorim, R. R. et al., 2006, "A Learning Design Ontology based on the IMS Specification", *Educational Technology & Society*, 9(1), pp. 38-57
- [2] Aswath, D. et al., 2005, "Boosting Item Keyword Search with Spreading Activation", *WI '05: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 704-707
- [3] Bauer, A.; Geisler, J., 2008, "Decision support for object recognition from multi-sensor data", In: Thoma, Klaus (Hrsg.); Fraunhofer Verbund für Verteidigungs- und Sicherheitsforschung: Future security: 3rd Security Research Conference Karlsruhe; 10th-11th September 2008: Congress Center Karlsruhe, Germany. Stuttgart: Fraunhofer IRB Verlag, pp. 321-326.
- [4] Bauer, A. & Herschelmann, O. Kamerman, G. W., 2010, "Exploiting context for assisted aerial image interpretation", *Electro-Optical Remote Sensing, Photonic Technologies, and Applications IV*, SPIE, 2010, 7835, 78350K+
- [5] Borst, T., 2006, „Ontologien zur semantischen Auszeichnung digitaler Lernmaterialien“, FernUniversität in Hagen Fachbereich Elektrotechnik & Informationstechnik, ed. (3). Available at: http://www.fernuni-hagen.de/imperia/md/content/fakultaetfuermathematikundinformatik/forschung/berichteit/forschungsbericht_3_2006.pdf (Retrieved February, 2011)
- [6] Cohen, P. R. and Kjeldsen, R., 1987, "Information retrieval by constrained spreading activation in semantic networks", *Inf. Process. Manage.*, 23(4), pp. 255-268
- [7] Collins, A. M. and Loftus, E. F., 1975, "A spreading-activation theory of semantic processing", *Psychological Review*, 82(6), pp. 407-428
- [8] Cootes, T. F. and Taylor, C. J., 2001, "Statistical Models of Appearance for Medical Image Analysis and Computer Vision", In *Proc. SPIE Medical Imaging*, pp. 236-248
- [9] Crestani, F., 1997, "Application of Spreading Activation Techniques in Information Retrieval", *Artificial Intelligence Review*, 11(6), pp. 453-482
- [10] Decker, S. et al., 1998, "Ontobroker: Ontology based Access to Distributed and Semi-Structured Information", *Database Semantics: Semantic Issues in Multimedia Systems*: Kluwer Academic Publisher, pp.351-369
- [11] Forberger, R. and Mareth, N., 2009, "Kopplung von Assistenz- und Ausbildungssystem im Bereich der Bildauswertung", In: A. Lichtenstein, ed. *Der Mensch im Mittelpunkt technischer Systeme: 8. Berliner Werkstatt Mensch-Maschine-Systeme; 7. bis 9. Oktober 2009*. Düsseldorf: VDI-Verlag (8. Berliner Werkstatt Mensch-Maschine-Systeme), pp. 444-450
- [12] Guha, R. et al., 2003, "Semantic search", *WWW '03: Proceedings of the 12th international conference on World Wide Web*. New York, NY, USA: ACM, pp.700-709
- [13] Henze, N. et al., 2004, "Reasoning and Ontologies for Personalized E-Learning in the Semantic Web", *Educational Technology & Society*, Volume 7 (Issue 4), pp. 82-97
- [14] Lichtenstein, A., ed., 2009, "Der Mensch im Mittelpunkt technischer Systeme: 8. Berliner Werkstatt Mensch-Maschine-Systeme", Düsseldorf: VDI-Verlag. 8. Berliner Werkstatt Mensch-Maschine-Systeme, 7. bis 9. Oktober 2009
- [15] Mareth, N.; Streicher, A.; Bauer, A. & Roller, W., 2010, „Context-aware retrieval of learning units“, *Proceedings of the IADIS International Conference e-Learning 2010*. Vol.1, 357-364
- [16] Miles, A. et al., 2005, "SKOS core: simple knowledge organisation for the web", *DCMI '05: Proceedings of the 2005 international conference on Dublin Core and metadata applications: Dublin Core Metadata Initiative*, pp. 1-9
- [17] Rocha, C. et al., 2004, "A hybrid approach for searching in the semantic web", *WWW '04: Proceedings of the 13th international conference on World Wide Web*. New York, NY, USA: ACM, pp. 374-383
- [18] Swertz, C., 2004, "Didaktisches Design: Ein Leitfaden für den Aufbau hypermedialer Lernsysteme mit der Web-Didaktik", W. Bertelsmann Verlag GmbH & Co. KG
- [19] Szentés, D. et al., 2008, "Computer-supported training for the interpretation of radar images", *7th European Conference on Synthetic Aperture Radar*, Friedrichshafen