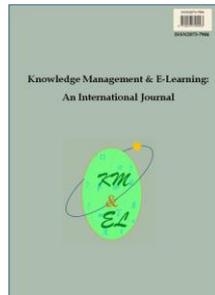


---

# **Knowledge Management & E-Learning: An International Journal**

---



ISSN 2073-7904

## **Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning**

**Nils Malzahn, Sabrina Ziebarth, H. Ulrich Hoppe**  
University of Duisburg-Essen, Duisburg, Germany

### **Recommended citation:**

Malzahn, N., Ziebarth, S., & Hoppe, H. U. (2013). Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning. *Knowledge Management & E-Learning*, 5(1), 84–103.

---

## **Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning**

---

Nils Malzahn\*

Faculty of Engineering  
University of Duisburg-Essen, Duisburg, Germany  
E-mail: malzahn@collide.info

Sabrina Ziebarth

Faculty of Engineering  
University of Duisburg-Essen, Duisburg, Germany  
E-mail: ziebarth@collide.info

H. Ulrich Hoppe

Faculty of Engineering  
University of Duisburg-Essen, Duisburg, Germany  
E-mail: hoppe@collide.info

\*Corresponding author

**Abstract:** Human resource managers are confronted with the problem that they have to fulfil the enterprise's competence needs either by developing their current staff or by recruiting new employees. In both cases decisions about who to select for the new position and more often which competences are crucial for the future success. This is especially true for highly dynamic industries like the IT industry. This article presents our work from the KoPIWA project in the Digital Economy. Our approach is based on a conceptual model that encompasses the market level, the social context and relations between competences. This model is the foundation for the ontology based decision support system for human resource managers presented in this article. To semi-automatically create and update the competence ontology methods from the areas data mining, social network analysis and information retrieval are employed. The results of these methods with regard to recruiting and learning processes are presented.

**Keywords:** Competence management; Decision support system; Conceptual model; Data mining; Ontology

**Biographical notes:** Nils Malzahn is a doctoral candidate at the department of computer science and applied cognitive sciences of the Faculty of Engineering, University of Duisburg-Essen. His research interests include architectures for ontology and agent-based support systems, especially for competence development and Social Network analysis. He is currently working on a project that investigates the benefits of Web 2.0 philosophy and architectures for adult training.

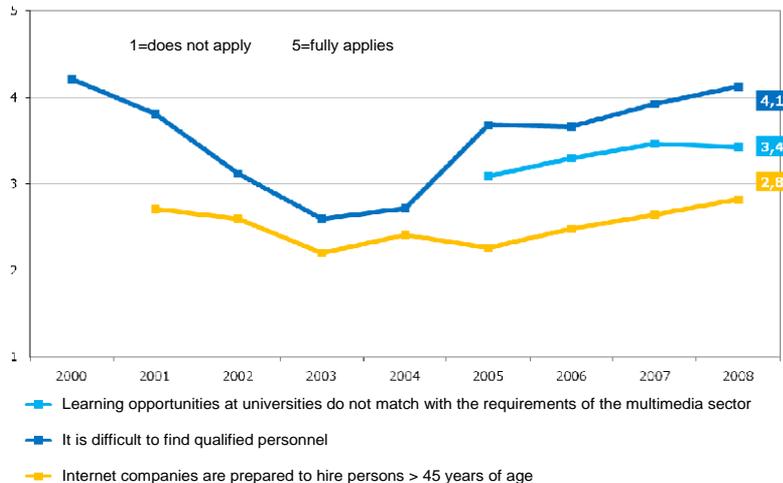
Sabrina Ziebarth is also a doctoral candidate at the department of computer science and applied cognitive sciences of Faculty of the Engineering,

University of Duisburg-Essen. Her research interests include data mining techniques for ontology learning and utilizing Serious Games for ontology and competence acquisition.

Dr. H. Ulrich Hoppe is Full Professor for collaborative learning in intelligent distributed environments (COLLIDE) in the department of computer science and applied cognitive science in the Faculty of Engineering at the University of Duisburg-Essen. He has been involved in multidisciplinary research in the areas of technology-enhanced learning, complex problem solving and learning, knowledge management and adult learning. His current research areas include cooperative learning and working environments, intelligent learning support systems and analysis and modelling of digital communities. More details can be found at <http://www.collide.info/>.

## 1. Introduction

According to figures of the online marketing association “Bundesverband digitale Wirtschaft” (see Fig. 1) for the important market segment of online marketing, the expected growth rate for 2009 was still 10%. This corresponds to about 3.500 new jobs based on average revenues. On this basis, the “war for talents” in the Digital Economy is ongoing (Hoppe, Malzahn, Mill, Zeini, & Hafkesbrink, 2010). In a nine year perspective, company representatives have expressed serious concerns about the potential of human resource acquisition (compare Fig. 1), especially from 2005 onwards (Erpenbeck & Michel, 2006).



**Fig. 1.** Human resource acquisition for the digital economy (Online Vermarkter Kreis, 2009)

Given the technical basis of the Digital Economy, this war for talents takes place in an area governed by high dynamics of technology, especially in the area of media convergence, which coincides with short innovation cycles and very limited half-life of knowledge and skills. These factors constitute an additional challenge for finding and hiring well trained personnel. It seems obvious that further growth rates of about 10-15% are only viable, if the core challenges of linking and pooling complementary knowledge

and skills to new business models are mastered adequately. (Hoppe, Malzahn, Mill, Zeini, & Hafkesbrink, 2010).

This article presents our findings from the KoPIWA project, funded by the German Federal Ministry of Education and Research. We will first introduce our basic understanding of competences and of process characteristics relevant for technology supported competence management. This defines our conceptual basis for developing the actual analysis and support tools, which are described in the following sections.

### *1.1. Notion of competence*

Following recent OECD studies, the development of competences has been identified as one important issue for the economic future of developed countries (DeSeCo, 2005). According to OECD individuals in modern societies need to acquire key competences to face the challenging demands of today's life. From this point of view, a competence is more than knowledge and skills. By the definition of OECD key competences must contribute to valued outcomes for society and individuals and help individuals meet demands in a wide variety of contexts. In this sense individual and societal success require individual and institutional competences as well as the application of individual competences to contribute to collective goals. Economic challenges often lead to situations where actors are confronted with changing contexts due to the dynamic development of the particular industry.

Some of the recent literature has distinguished the notions of 'competence' and 'competency'. In this sense, competency would be seen as the ability to act or perform, or, in other words, the set of competencies would form a "behavioural repertoire" (Woodruffe, 1993) (Kurz & Bartram, 2002), whereas competence would be defined as a requirement for a certain job or task. This distinction leads to seeing competency as a deeper and more valid concept in that it is related to actual behaviour. However, we will argue that the matching of job requirements and offers of workforce does often work without validating or measuring behavioural dispositions. Thus, the mechanisms that we claim to be practically relevant are of a more declarative or "shallow" nature. Accordingly, we will stick to using the term "competence".

While the international discussion on the development of competences is often interwoven with research on knowledge management (Davenport & Prusak, 1998), the scientific discourse in Germany has been particularly centred on a broad and differentiated notion of competences, including social skills. In this discussion, competences have been characterized as self-organization dispositions from a subject-related "entire person" perspective (Erpenbeck & Michel, 2006). Considering the highly dynamic nature of the IT and digital media field, which forms the background of the digital economy, the focus on subject-centred and autonomous self-organization of dispositions is indeed more adequate than a traditional perspective on externally induced qualification and training. In a previous study of IT freelancers, we have characterized personal competence development as "Selbstaktualisierung" (autonomous updating of oneself), i.e. the autonomous self-organized extension of skills and knowledge (Shire, Borchert, & Hoppe, 2007). Workers in the IT and digital media sector have to develop their competences in relation to their social networks as an activity of lifelong-learning. (Hoppe, Malzahn, Mill, Zeini, & Hafkesbrink, 2010). The synergy between social-skills and professional expertise captured by the concept of self-organized competence development within social networks goes beyond traditional assumptions about externally organized training. In this sense, (Erpenbeck & Sauter, 2007) suggest to complement competence assessment tools such as KODE with e-portfolios. Another

relevant example of a "Personal Competence Manager" (Kew, 2007) has been developed in the European TENCompetence project.

### *1.2. Competence management as a negotiation process*

Human resource management is focused on identifying and developing competences on an organizational level. Within this context, we will particularly look at the process of recruiting. There is a widely accepted tendency to define competence management in recruiting and training as a rational decision making process with competences as measurable variables. Thus, the HR-XML consortium defines a competence as "a specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic, such as attitude, behaviour, physical ability, which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context" (HR-XML Consortium, 2003).

For both practical and theoretical reasons it is questionable if this strong assumption about measurability can actually be defended. In the tradition of psychology, Weinert (2001) challenges this claim quite strongly: „An exhaustive definition of competence would have to include all the intellectual abilities, content-specific knowledge, skills, strategies, metacognitions and action routines that contribute to learning, problem solving and a variety of achievements. One would be confronted with a problem not yet solved in the 100-year history of scientific psychology: a complementary classification and performance-specific integration of ability and knowledge. There is neither a theoretical nor a practical solution to this problem at this time." Of course, psychology has developed scales and measurement procedures for certain behavioural dispositions, certain cognitive abilities or skills. But Weinert's point is that we are far from possessing a complete set of such instruments to assess all types of competences relevant to HR management.

Disregarding this theoretical controversy, we can state with (Lindblom, 1959) that the everyday practice of human resource management is rather characterized by following the "science of muddling through". The notion of competences used by HR managers use is fuzzy and often inconsistent (see next Section). Following (Strauss, 1993), the acquisition of competencies for an enterprise can be viewed as part of its inherent processual ordering mechanisms.

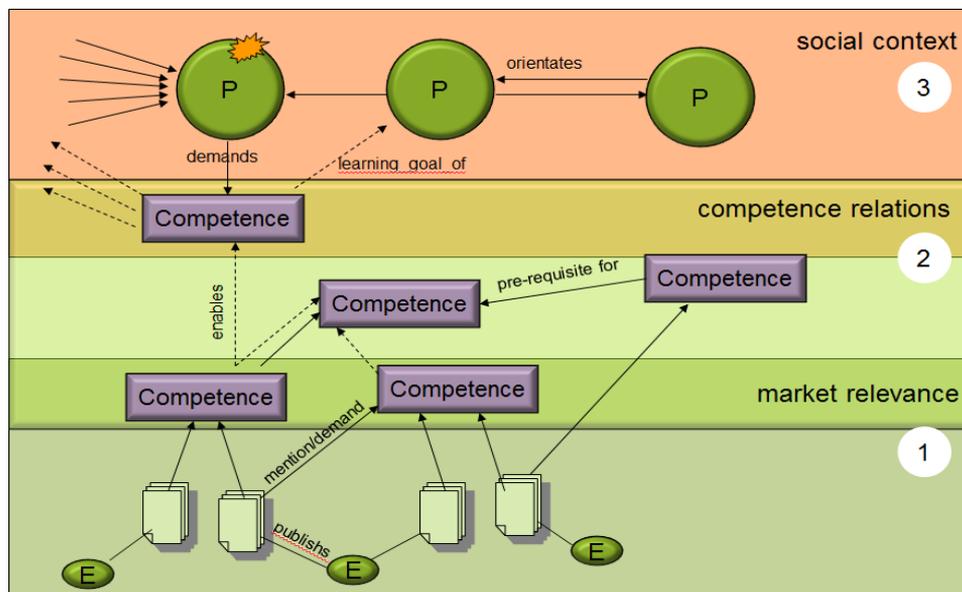
We believe that competence management can be reasonably studied and understood without making strong assumptions about measurability and operationalization. As an alternative, we propose a view, which conceives competence management and especially recruiting as a negotiation process. This negotiation process involves certain actors (such as HR staff, applicants or job candidates), and it is substantiated by means and acts of communication such as job offers, job interviews, or training procedures. We propose to see these communication activities as elements of a language game in the sense of (Wittgenstein, 1953). Using this term we intend to point out the communicative character of competence management as well as the claim that the meaning of the communication of competence requirements and offers is not context free. In such language games, competences are not identified by "objective" definitions but pragmatically by "family resemblance" (Wittgenstein, 1953).

Pragmatically, we see the negotiation or language game interpretation of recruiting and competence management as an analytic reference frame that avoids the high claims of measurability and operationalization. We do not have to follow Weinert in dismantling these high claims as being illusory. It is already clear from a practical

perspective that most everyday recruiting procedures are not based on validated measurements of competences (Erpenbeck, 2009). The argument about "regulation by expectation" also shows that a language game would not just render arbitrary decisions, but that there is a form of a subjective evaluation on the part of both negotiating parties.

## 2. An ontology-based conceptual model for competence development

The general approach towards competence management is inspired by a conceptual model (Malzahn, Urspruch, Zeini, & Hoppe, 2007). The model inter-relates three layers or levels, namely the market level (1), the individual competence level (2) and the social network level (3) (see Fig. 2). The market level is defined by current skill demands and trends, possibly extrapolated to a near future. We have used job offers to assess the given market needs (see Sub-section 3.2). On the individual competence level, the skill nodes from the market level are semi-automatically combined and inter-connected with each other based on co-occurrences (see Sub-section 3.3ff.) and with deeper level competences. The social network level takes into account that professionals are part of a social network, comprised of peer relations to colleagues as well relations to companies. This network partly represents the actor's social capital.



**Fig. 2.** Conceptual model

The evolution of individual and organizational competences should target new market demands, yet counting as much as possible on existing knowledge and skills and the given social network – especially in the case of freelancers. Certain new demands may be (too) costly in terms of additional learning effort, and certain changes of professional fields may put the social capital at risk. In this sense, the model can be interpreted in terms of costs and expected benefits. It is focused on the individual for a start, but it can also be employed in a company to support individual career planning and, in a summary view, human resource management. In the following subsections the three levels are explained in more detail.

### *2.1. Market level*

To identify competences on the market level we have extracted skill descriptions from job offers and publicly available personal profiles. From our analysis of job offers (Ziebarth, Malzahn, & Hoppe, 2009) in the IT domain, we can tell that most offers in this domain target surface level competences that can immediately be applied to present tasks and the currently used tools. Unfortunately these competences do not address the deeper understanding of a domain which may allow the individual to keep track with future changes. Thus surface level competences tend to become obsolete quickly, whereas deep level competences may be of long term value, because they enable the individuals to transfer their knowledge to new problems and tasks.

The competences are weighted based on the number of occurrences in job offers per time slice, i.e. if these skills are mentioned in more recent job offers their weight is higher. In this sense a first, simple trend analysis may be employed in terms of increasing demand or decreasing demand. Furthermore this level may be the foundation of a dynamic ontology (see Section 3).

### *2.2. Competence level*

The competence level is based on an ontology of competences, which has been partly built and updated by the methods described in Section 3 and partly engineered by domain experts. The (weighted) skill nodes from the market level are combined and set into relation with each other and the deeper level competences. To refine the initial assessment of the competences two relations are introduced on the competence level. One relation describes the individually adapted, estimated effort that needs to be spent to acquire a competence, given that the competence on the starting point of the arrow has already been learned. The second relation describes which surface-level competences are encompassed or backed-up by a particular deeper level competence. Thus, the ontology makes an explicit differentiation between surface level competences (for example skills) and deep level “generic” competences. Furthermore, we assume that deep level competences allow the individual to learn a corresponding bundle of surface level competences, because of the knowledge about the underlying concepts.

Within this representation we can define learning paths through the ontology. A learning path is the set of competences in a given order that should be acquired by an individual, either by the individual’s desire or to meet a pre-requisite for another desired competence. The direction of the relations in the ontology embosses a (partial) order onto the sequence of learning, given a starting set of competences that has already been acquired (current competence profile) and the target set of competences (target competence profile).

The information based on the conceptual model’s level one and two allow for an assessment of a specific competence based on simple graph characteristics like in-degree (amount of incoming edges) and out-degree (amount of outgoing edges) of a competence node in the combined competence job-offer graph. First, a competence may be evaluated solely based on the competence level. This results in a classification of the competence in the portfolio shown in Table 1. Depending on the current (personal) situation and (external) requirements, one sector of the portfolio maybe preferred. In a running project, the acquisition of an island competence for example may be more reasonable than an enabler competence to meet the current project’s goals, although the particular competence does not directly help to acquire new projects.

**Table 1**

Competence portfolio classified by in- and out-degree concerning potential learning paths

		In-degree	
		low	high
Out-degree	low	Island competence	Expert competence
		Pre-requisite for few other competences, but needs little prior knowledge	Pre-requisite for few other competences, but needs extensive prior knowledge
		Project specific qualification	Useful for further qualification
	high	Enabler competence	Generalist's competence
Pre-requisite for a lot of other competences and needs little prior knowledge		Pre-requisite for a lot of other competences, but needs a wide spectrum of prior knowledge	
Opens up new domains		Potential managerial competence	

**Table 2**

Assessment of competence based on offer and demand

		In-degree	
		low	high
Out-degree	low	Specialist's competence	Market trend
		Low demand, but also few people offering it.	High (increasing) demand and currently only a few persons offer it.
		e.g. Energy plant software programming	e.g. Dot net, SAAS
		Recommendation: may be acquired, if fitting to customers and current profile	Useful for further qualification
	high	Obsolete competence	Basic competence
		Low (decreasing) demand and a lot of people offer it.	Nearly everyone has it and it is often needed
e.g. Cobol programming		e.g. object-oriented programming	
	Recommendation: do not learn	Recommendation: To be learned with others	

Second, the competences may be assessed concerning offer and demand, which is basically also evaluated by in- and out-degree of the combined market-level and competence-level network (see Table 2). It offers a more objective way of judging the estimated usefulness of a newly acquired competence. Exceptions are specialist's competences, because their usefulness is heavily dependent on the prior knowledge of the learner. If the effort to learn such a competence is low, because the prior knowledge allows for a steep learning curve, it may be worth the costs. Since such competencies are usually acquired by persons rooted in the special domain of interest, it is also difficult to enter such a domain without the experience made in the domain. Thus, this competence should not be acquired by others. The other three quadrants are more straight-forward: if there is a high demand, one should be able to fulfil it, otherwise look for another competence to learn.

2.3. Social context level

The discussion of the usefulness of specialist’s competences already includes some considerations about the social context of the learner. The social network level explicitly takes into account that professionals and enterprises are part of a social network that represents at least a part of their social capital. Thus, changes caused by decisions for new learning and career opportunities, may have impact on the social capital. This may happen for example if the professionals try to orientate themselves towards new technologies which are not part of the portfolio of their current employers, customers or more general of the team they are currently working in.

The conceptual model represents this context by modeling the relations between social actors, especially persons, using three different relations. The first one was already mentioned in the market level. It indicates that an actor (person/enterprise) has a demand for a specific competence. The second one represents the known wish to acquire a specific competence. This may be known for example because of a discussion. The last one expresses that a person uses another actor for orientation purposes. This may be the case either because the other actor is a business partner or because it is a person who is known to be an evangelist or a successful visionary. The later type of person is a trusted authority (Harrer, Malzahn, Zeini, & Hoppe, 2007). These persons are specifically marked in the ontology, because their judgments should be particularly considered, when deciding what to learn next.

2.4. Discussion of the conceptual model

The conceptual model can be used as the basis for a decision support system for competence management. As explained above relevant job profiles can be modeled in the ontology, which enables the support system to assist the professionals in their career planning. Together with modelled relationships between the competences, a path from the professional’s current profile to the targeted position can be inferred and suggested (see Fig. 3).

Since there will usually be more than one path from the current set of competences to the targeted set of competences, the other two levels of the comprehensive model are considered while recommending a certain path. The market level increases or decreases the importance of particular surface skills. This is important to stay employable. The social network level increases or decreases the importance of the competences that are held or expected by the network buddies (either peers or organizations) of the individual IT worker. In the end the support system will present a set of to-be-acquired competences based on an overall ranking value derived from the learning effort needed, the gain or loss of social capital and the market demands.

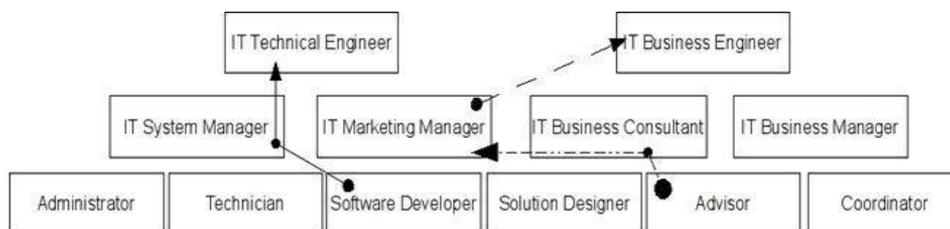


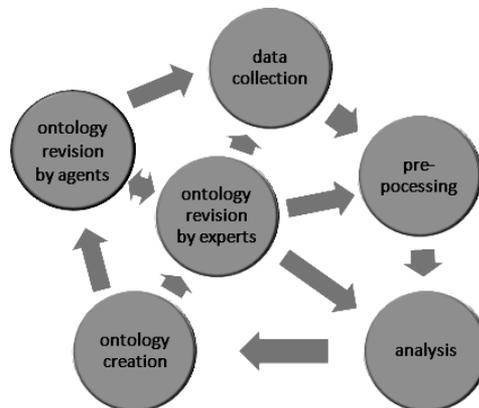
Fig. 3. Example for alternative trajectories in individual professional development

A small scale study with students of computer science showed that the proposed model is applicable to real cases even with a very simple formula (simple addition of estimated gains minus estimated costs) for the overall evaluation of the possible target competences. (Schröder, 2007)

As the model is based on data that is frequently updated with the market demand, the evolving ontology and the changes in the social network it is capable of capturing the dynamics of the specific branch of trade. Obviously having an up-to-date and appropriately expressive ontology is crucial to the success of the model. In the following sections we will describe in detail how to derive and maintain such an ontology based on publicly available data.

### 3. Semi-automatic creation and maintenance of competence ontologies

In highly dynamic sectors like the digital media industry new concepts are frequently coming up whereas others are considered obsolete at a high pace. To manually create and maintain an up-to-date competence ontology in such a dynamic field is highly time consuming, expensive and thus almost not feasible (Hepp, 2006). Therefore, we propose a semi-automatic approach to the creation, refinement and maintenance of competence ontologies.



**Fig. 4.** Process of ontology creation, refinement and maintenance

The overall process consists of six sub-processes (see Fig. 4): Artefacts containing information about competences are automatically collected (data collection), pre-processed and analyzed using methods from data mining, information retrieval and social network analysis. The resulting raw competence ontology can be refined manually by domain experts and agents using ontology reconciliation. The refined competence ontology can be injected into the overall process again to tap into new data sources and enhance the analysis of the continuously collected new data.

The extraction of the raw competence ontology is conducted by a multi-agent framework, which is described in the next section.

#### 3.1. Multi-agent framework for ontology creation/maintenance

Multi-agent frameworks have a tradition in being used for Natural Language Processing (NLP), since they support functional decomposition of NLP problems into sub-problems,

which can be solved by (regarding representation and reasoning technique) specialized, independent, and parallel running agents (Erman & Lesser, 1980). Single sub-problems can be allocated to several agents with different viewpoints integrating different theories of cognitive processing (Fum, Guida, & Tasso, 1988). Furthermore, the independence and redundancy of the agents gives the system a high degree of robustness. Therefore, our agent framework (see Fig. 5) uses a blackboard architecture, like (Erman & Lesser, 1980), implemented by a Tuple Space (Weinbrenner, Giemza, & Hoppe, 2007).

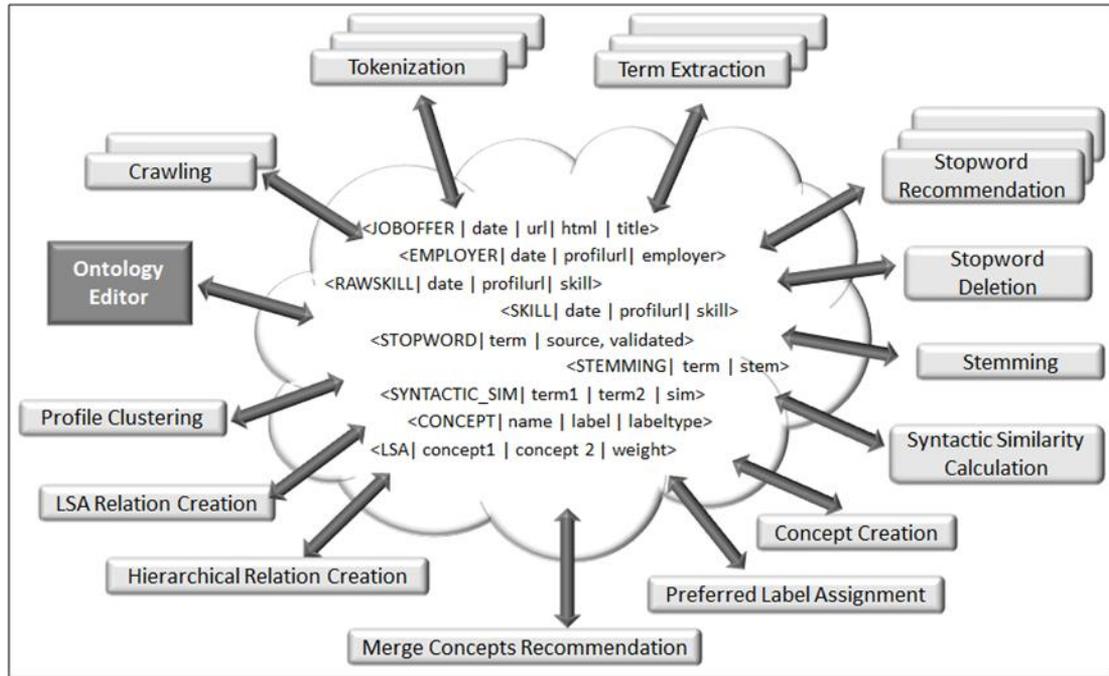


Fig. 5. Multi-agent framework (partial view)

There are several agents with tasks ranging from data collection, such as crawling, pre-processing to the analysis and clustering of the data. The single agents as well as some of their results are presented in the next sections. While the data and the retrieved competence ontology are in German, the presented results are (mainly) translated to English for better understanding.

### 3.2. Data and pre-processing

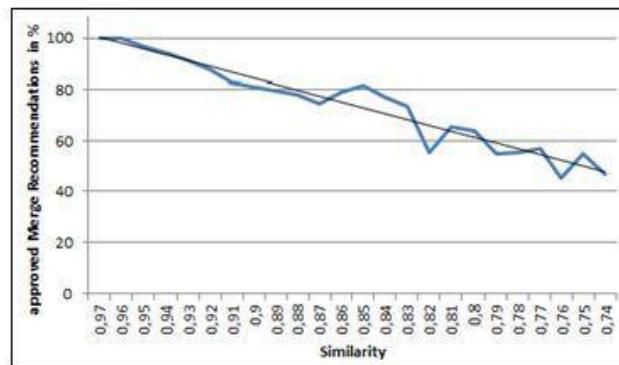
Since most job offers in the IT and media sector are published on the internet either on company homepages or in online job portals, they provide an easily accessible source of information about job profiles, their titles and the requested competencies. Thus, the data collection agents retrieve current job offers as well as user profiles. In our case study we only considered German job offers and user profiles of the IT industry. Our initial data set consisted of 23.536 semi-structured person profiles and around 62.000 free text job offers.

Since the available person profiles are relatively well-structured, they can be used to identify competence labels rather easily, compared to the free text job offers. Thus, the person profiles are used for creating a competence dictionary, which is applied to the job

offers. The tokenization and extraction agents try to extract specific information like job titles, competence labels, employers and organizations utilizing the person profiles' structure that contains comma separated free-text terms classified in the categories skills, employers, and organizations. These agents identified a preliminary set of 15.149 job titles, 143.300 competence labels, 87.744 employers and 31.913 organizations.

Since not all of the extracted competence labels really describe competences, they are filtered using pre-defined stop word lists, information from Germa-Net and Wortschatz complemented by statistical measures from text mining, such as identifying stopwords by their (relative) frequency (Fox, 1989). The detected stop words are tagged as possible stop words in the Tuple Space. While the semantic approach is quite accurate (around 85% of the GermaNet and 76% of the Wortschatz stop words were approved by experts), the statistical approach is not feasible (only about 16% of the stop words were approved). Since the skill lists contain barely general stop words like pronouns or conjunctions, there are few stop words that have a sufficiently high occurrence, whereas the important words have a relatively high occurrence.

A competence can be described by different labels, resulting from different word forms, usage of technical terms in different languages or due to spelling errors. For later analysis, similar labels are combined to concepts. Since conflation of word forms can enhance the quality of information retrieval significantly (Frakes, 1984), the Porter Stemmer (Porter, 1980) is used to merge concepts with the same stem assigning the original labels to it. Furthermore, the syntactical similarity between the skill labels is calculated based on the Levenshtein distance (Levenshtein, 1966). Words with a high syntactical similarity are good merging candidates (see Fig. 6). Nevertheless, they are not merged automatically, since there are lots of words, which are syntactically and even semantically very similar, but describe not exactly the same competence, for example "Reporting" and "eReporting" or "Java J2EE" and "Java J2SE". Furthermore, we want to distinguish between different software versions, such as "Microsoft Office 2000" and "Microsoft Office 2010", as those versions might differ significantly.



**Fig. 6.** Approved merge recommendations

### 3.3. Identification of hierarchical competence relations

For the detection of hierarchical relations we use heuristics based on the analysis of n-grams and lexico-syntactic patterns. Labels of competences differing only in the first or last term, such as "Java" and "Java 3D", are considered to have a hierarchical relation with the shorter word being more general (for example "Java" is broader than "Java 3D").

About 77% of the identified word pairs were approved as hierarchical relations, 21% merged and less than 2% were fully rejected. Additionally we applied Hearst’s six lexico-semantic patterns for detecting hierarchical relations in a free text (Hearst, 1992) to the free text job offers. Table 3 gives an example for derived hierarchical relations.

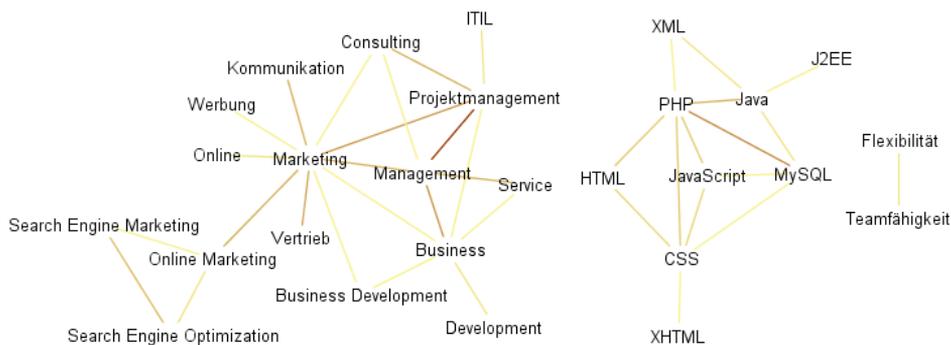
**Table 3**  
Examples for matching lexico-syntactic patterns

Example	Relations
... development tools <i>such as</i> Eclipse, Ant, Maven <i>and</i> SVN ...	Development tools > Eclipse
	Development tools > Ant
	Development tools > Maven
	Development tools > SVN

3.4. Identification of associative competence relations and rules

In addition to the extraction of hierarchical relations, associative relations based on frequent co-occurrences are retrieved. These are identified by using the apriori algorithm (Agrawal, Mannila, Srikant, Toivonen, & Verkamo, 1996) to efficiently calculate large item sets containing two concepts, i.e. all combinations of competence concepts, which occur with a minimal frequency in the corpus.

The network extracted from around 23.500 person profiles using a minimal frequency of 0.01 is displayed in Fig. 7. There are three cohesive clusters of competences, presenting different aspects of the digital economy. The biggest cluster contains competence from the field of management and marketing, which are rather generic and can be applied to different industries. Specific for the digital economy is only the triad consisting of online marketing, search engine marketing and search engine optimization. The second cluster contains technologies from the field of web development and the last one the two soft skills flexibility and team work. The web-development cluster can already be interpreted as a profile of a typical web developer. This information can be used to identify neighbouring industries or job profiles where talents might also be found.

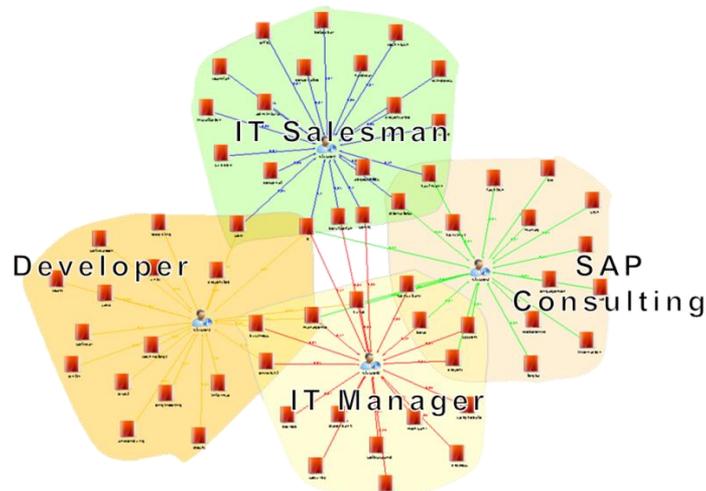


**Fig. 7.** Associative relations based on person profiles (darker relations are stronger)

Using association rule techniques competence rules have been extracted from the profiles. Depending on the data, the meaning of the rules differs: Rules extracted from person profiles reveal typical sets of competences of real employees, while rules based on job offers state demands for a fictional person. Thus, the rules help to specify a profile of a talent.

### 3.5. Identification of job profiles

Clustering techniques are used in data mining to decompose example sets into clusters of similar examples (Witten & Frank, 2005). Clustering job offers has proved the existence of clusters of similar jobs in the corpus of job offers and given an insight into the competences discriminating the clusters (Ziebarth, Malzahn, & Hoppe, 2009). Additionally emerging new profiles from combinations of existing ones at intersections of existing profiles may be found (see Fig. 7). The emerging job profiles may be seen as the profiles of the talents of the future, as they indicate a shift in needed competences. If a recruiter can take into account the competence profiles that are currently emerging, there is a head start in recruiting a talent to be warred for.



**Fig. 7.** Overlapping clusters of job profile competencies

## 4. Support of profiling, matching and competence development

The semi-automatically created competence ontology is the foundation for a tool, which supports the creation, matching and development of competence profiles (see Fig. 8). Competence profiles can be created automatically by loading an available job announcement or personal data sheet, or manually by selecting competences found by keyword search or browsing the competence ontology which is visualized as a tree. Thus, on the one hand, the ontology is used as domain specific vocabulary. On the other hand, it is used for matching competence profiles. Furthermore, users can assign weights to the importance of the competences (low, middle, high) in their profiles.

The matching of competence profiles has multiple applications: Human resource managers can test if their job offers are similar enough to other ones / job profiles to be found by applicants and if they are specific enough to distinguish them. Furthermore, they can test which kind of person could apply to the job offer based on the matching person profiles. By comparing similar profiles they might get some inspiration for improving their own ones. Employees can test, which kind of jobs match their profiles and thus search more purposefully since they are provided with specific job titles. They may also detect competences they forgot or considered unimportant, which can be applied in their professional life or which they want to learn to enhance their own profiles.

After creating a profile, users are provided with feedback. The distribution of the competence types (professional, methodical, social, personal) is visualized. Since professional competences discriminate competences better than other competence types (Ziebarth, Malzahn, & Hoppe, 2009), typical profiles are dominated by them, resulting in “pacman” like charts (see Fig. 8). Thus, while the lack of each of the competence types is communicated to the user, deficits in the professional competences are highlighted in particular. Furthermore, typical extensions (based on the association rules) and job titles for the profile are suggested.

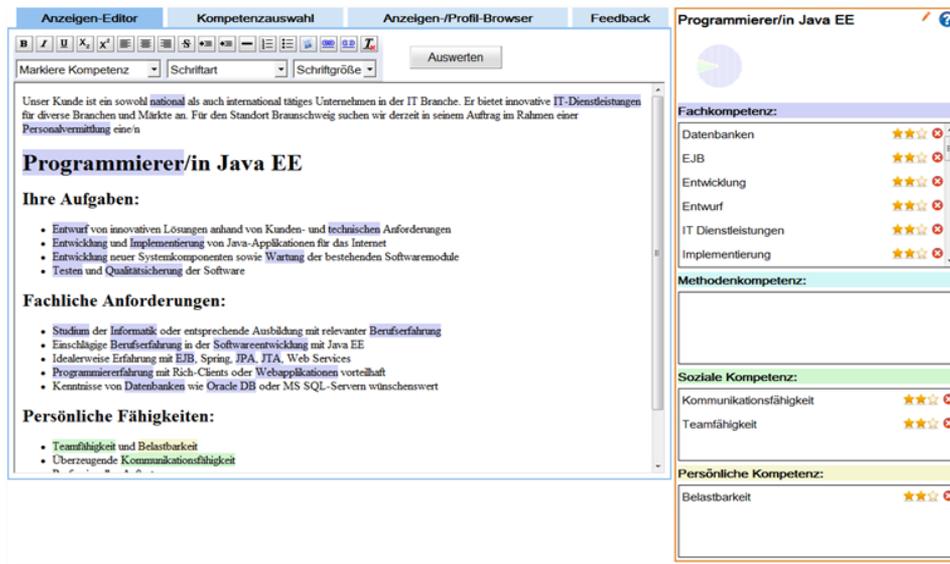


Fig. 8. KoPIWA profiling and matching tool

The semantic information stored in the competence ontology is used in the matching process for competence profiles. One approach is to enhance the well-known vector space model (Salton, Wong, & Yang, 1975) with term/query expansion by including not only the competences themselves, but also their parent, sibling or children competences. Another approach evaluates the semantic similarity of the competences contained in the profiles (Biesalski & Abecker, 2006). There are several methods for measuring the similarity of concepts in semantic nets: Rada, Mili, Bicknell, and Blettner (1989) showed that the shortest path (minimal amount of edges) between two concepts in an inheritance-hierarchy is a good measure for a conceptual distance. Resnik (1995) suggested using the concepts’ information content, which is defined by the probability of encountering an instance of the concept in a given corpus.

Since human similarity assessment is context dependent (Tversky, 1977) or (Medin, Goldstone, & Gentner, 1993), to synthesize an optimum measure for the ontology based comparison of competence profiles, we conducted an experiment comparing user and system assessment (Ziebarth, Malzahn, Daems, & Hoppe, 2010). The subjects selected - in addition to professional competences - many “soft skills” for their profiles, but focused on the professional competences when assessing the results. Thus, measures focussing on the depth of a competence in the ontology are better related to the user expectation. The tendency of the subjects to focus on professional competences is plausible, since they represent the distinctive factors of job offers while “soft skills” are more generic (Ziebarth, Malzahn, & Hoppe, 2009). Although the results are not

statistically firm, they hint at issues to be taken into account for competence profile matching: “soft skills are considered important, but they not perceived as relevant for job offer selection as discriminating professional competences.” Thus, to improve the probability that a high potential talent applies for a job because of a job offer, job offers should focus on discriminating and specific professional competencies instead of common sense soft skills.

## 5. Related work

There are several approaches, which use competence ontologies as shared domain model (mainly for profiling, gap analysis and profile matching) in competence-based management, for example Zelewski (2005) uses competence ontologies for cooperative knowledge exchange in engineering networks. Kunzmann and Schmidt (2006) use them for health care training planning. Draganidis, Chamopoulou, and Mentzas (2006) apply competence ontologies to a wide range of HR managements tasks as well as to generate learning path including learning objects. Dorn, Naz, and Pichlmair (2007) use competence ontology for e-recruitment and a meta-search engine to search for appropriate jobs for students. Biesalski and Abecker (2006) try to find the optimal allocation of project positions based on the available human resources. Hefke and Stojanovic (2004) use a competence ontology to create ad hoc teams complying with a given project profile.

The ontologies used in competence-based management are typically handcrafted by collecting domain knowledge (competences used in a certain branch) from domain experts and aggregating this knowledge by modeling experts and thus contain only few instances, for example (Zelewski, 2005), (Kunzmann & Schmidt, 2006), (Dorn, Naz, & Pichlmair, 2007), (Sure, Maedche, & Staab, 2000), and (Ley et al., 2008).

There is an increasing number of approaches that use data mining techniques to assist human resource management (Piazza & Strohmeier, 2011), but there are only few not very elaborated approaches, which try to extract competences and their relations from artifacts. Rodrigues, Oliveira, and de Souza (2004) use text mining methods to identify competences of researchers based on their publications. Zhu, Goncalves, Uren, Motta, and Pacheco (2005) use named entity recognition to extract persons, organizations, projects and research topics from business documents. The identified entities are linked by an analysis of their co-occurrence. All links of persons to other entities are interpreted as their competences (also persons, organizations and projects), which is a rather broad interpretation of competences. Apart from these domain specific approaches, there are much more elaborated domain independent ones in the context of ontology learning, for example (Maedche & Staab, 2001) or (Fortuna, Grobelnik, & Mladenic, 2005).

Generally, a lot of work has been performed regarding the inclusion of ontologies to enhance the quality of matching, for example Castellás, Fernández, and Vallet (2007) provide an adaption of the VSM for ontology-based information retrieval, Li, Bandar, and McLean (2003) compare measures for semantic similarity of words and Ehrig, Haase, Hefke, and Stojanovic (2004) provide a comprehensive framework for measuring similarity between entities within an ontology. But there is only few work like Biesalski and Abecker (2006) or Hefke and Stojanovic (2004), describing how the structure of the competence ontology can be considered for matching competence profiles; although, they provide no evaluation of the matching quality.

## **6. Discussion**

Starting with the definition of a pragmatic competence concept that is based on the assumption that people are rational concerning their claims about their individual competence, but uncertain about their own development needs, we introduced a conceptual model. On the basis of this model a decision support system for human resource managers has been created, that supports them in recruitment tasks as well as competence development, especially career planning.

One foundation of the decision support system is a semi-automatically maintained ontology, which allows for trend detection with respect to competence importance and emerging job profiles. Thus, it enables human resource managers to react timely to future requirements.

As the model is based on data that is frequently updated like the market demands, the evolving ontology and the changes in the social network, it is capable of capturing the dynamics of the specific branch of trade. If it is used and maintained over a longer period of time it is also possible to capture the dynamics of the personal profile. This may lead to an even better decision support, because the zone of proximal development (Vygotski, 1978) can be determined more accurately.

Nonetheless, the system is not capable of functioning fully automatically. Domain experts have to evaluate newly found competences and relate them to others. For this reason we developed an ontology editor named LOWL (Light-Weight OWL) that is capable to cope with the big amount of data that is easily gathered from public data sources by working directly on the SQLSpaces. Still, the amount of data as such makes it difficult to distinguish precisely between competences and “noise”, but the quality is increasing steadily with each human intervention. The final decision if a term is an upcoming competence or a just a new word for a well-known one has to be made by a human expert.

Another limitation of the presented approach is that the conceptual model demands for a reasonably huge part of the users’ social context. This will always be a human effort and might present a cold start problem, because the system will be more efficient the more people take part in it.

At the moment the system is designed as a close one, inter-enterprise. Of course it would be possible to think of a publicly available system. In this case privacy issues have to be observed, especially on the social context level. There has to be a trade-off between the transparency of the recommended learning paths and the visibility of the competence profiles or wishes in the social context, since the users will want to understand why a particular competence has been proposed, but they might not want to show that they are interested in a specific competence, for example because it is a confession of “not knowing”.

In summary, the presented system provides a means to get an advantage in the “war for talents”. It allows for matching job profiles with employer’s profiles and since the needed talents may just not be available it enables the human resource managers to develop the needed talents from their staff by supporting the career planning at an early stage. Thus, the talents may be available when they are needed to thrive on innovation and staying capable of competing.

While the prototype of the system has not been directly taken up by the industry there are set follow-up projects that are based on the presented concepts. Currently the German economy is trying to fill the upcoming gap of skilled workers with their current

staff. One of the main approaches is to uncover the potential of the existing staff by considering their “secondary” competences, which are acquired during leisure times, for example while pursuing hobbies. The ontology-based approach presented here is adapted to relate “hobby” competences to professional ones. Another approach is to exploit the experience of aging staff at other positions to prolong their working life. Accordingly, the career planning approach of the conceptual model is systematically applied.

### Acknowledgements

This research was supported by the German Ministry of Education and Research by funding the project KoPIWA (No. 01FM07067-72) on competence development in open innovation networks in the IT and media sector.

### References

- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. (1996). Fast discovery of association rules. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 307–328). AAI Press / MIT Press.
- Biesalski, E., & Abecker, A. (2006). Skill-profile matching with similarity measures. *Proceedings of 8th International Conference on Enterprise Information Systems (ICEIS)* (pp. 210–218). Paphos, Cyprus: Springer.
- Castellas, M., Fernández, M., & Vallet, D. (2007). An adaptation of the vector space model for ontology-based information retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 19(2), 261–272.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Cambridge: Harvard Business School Press.
- DeSeCo. (2005). *Definition and selection of competencies: Theoretical and conceptual foundations*. OECD.
- Dorn, J., Naz, T., & Pichlmair, M. (2007). Ontology development for human resource management. *Proceedings of the 2007 International Conference on Knowledge Management (I-KNOW '07)* (pp. 109–120). Graz, Austria.
- Draganidis, F., Chamopoulou, P., & Mentzas, G. (2006). An ontology based tool for competency management and learning paths. *Proceedings of the International Conference of Knowledge Management (I-Know '06)*. Graz.
- Ehrig, M., Haase, M., Hefke, M., & Stojanovic, N. (2004). Similarity for ontologies - A comprehensive framework. *Proceedings of the Workshop on Enterprise Modeling and Ontology: Ingredients for Interoperability*. PAKM.
- Erman, L., & Lesser, V. R. (1980). The HEARSAY-II speech understanding system: Integrating knowledge to resolve uncertainty. *Computing Surveys*, 12, 213–253.
- Erpenbeck, J. (2009). Kompetente kompetenzerfassung in beruf und betrieb. In E. Severing, & H. D. Münk (Eds.), *Theorie und Praxis der Kompetenzfeststellung im Betrieb - Status Quo und Entwicklungsbedarf*. Bundesinstitut für Berufsbildung.
- Erpenbeck, J., & Michel, L. (2006). Competency-based quality securing of e-learning (CQ-E). In U.-D. Ehlers, & J. M. Pawlowski (Eds.), *Handbook on Quality and Standardisation in E-Learning* (pp. 125–141). Berlin: Springer.
- Erpenbeck, J., & Sauter, W. (2007). *Kompetenzentwicklung im netz*. Cologne: Luchterhand Verlag.

- Fortuna, B., Grobelnik, M., & Mladenic, D. (2005). Semi-automatic construction of topic ontology. *Proceedings of the ECML/PKDD Workshop on Knowledge Discovery for Ontologies*.
- Fox, C. (1989). A stop list for general text. *ACM SIGIR Forum*, 24(1-2), 19–21.
- Frakes, W. (1984). Term con. *Proceedings of the 7th annual international ACM SIGIR conference* (pp. 383–389).
- Fum, D., Guida, G., & Tasso, C. (1988). A distributed multi-agent architecture for natural language processing. *Proceedings of the 12th conference on Computational linguistics* (pp. 812–814). Budapest, Hungary.
- Harrer, A., Malzahn, N., Zeini, S., & Hoppe, H. U. (2007). Combining social network analysis with semantic relations to support the evolution of a scientific community. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, Minds, and Society - The Computer Supported Collaborative Learning (CSCL) Conference* (pp. 267–276).
- Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th conference on Computational linguistics* (pp. 539–545). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Hefke, M., & Stojanovic, L. (2004). An ontology-based approach for competence bundling and composition of ad-hoc teams in an organisation. *Proceedings of the International Conference of Knowledge Management (I-KNOW '04)*.
- Hepp, M. (2006). Possible ontologies: How reality constrains building relevant ontologies. *IEEE Internet Computing*, 11(1), 90–96.
- Hoppe, H. U., Malzahn, N., Mill, U., Zeini, S., & Hafkesbrink, J. (2010). Negotiating competences in recruiting for highly dynamic work contexts. In J. Hafkesbrink, H. U. Hoppe, & J. Schlichter (Eds.), *Competence Management for Open Innovation* (pp. 53–67). Cologne: EUL-Verlag.
- HR-XML Consortium. (2003). *Competences (Measurable Characteristics)*. Retrieved from [http://ns.hr-xml.org/2\\_0/HR-XML-2\\_0/CPO/Competencies.pdf](http://ns.hr-xml.org/2_0/HR-XML-2_0/CPO/Competencies.pdf).
- Kew, C. (2007). The TENCompetence personal competence manager. *Proceedings of EC-TEL Conference*.
- Kunzmann, C., & Schmidt, A. (2006). Ontology-based competence management for healthcare training planning: A case study. *Proceedings of the International Conference on Knowledge Management (I-KNOW '06)*. Graz.
- Kurz, R., & Bartram, D. (2002). Competency and individual performance: modelling the world of work. In I. Robertson, M. Callinam, & D. Bartram (Eds.), *Organizational Effectiveness: The Role of Psychology* (pp. 227–255). Wiley & Sons Ltd.
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- Ley, T., Ulbrich, A., Scheir, P., Lindstaedt, S. N., Kump, B., & Albert, D. (2008). Modeling competences for supporting work-integrated learning in knowledge work. *Journal of Knowledge Management*, 12(6), 31–47.
- Li, Y., Bandar, A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 871–882.
- Lindblom, C. E. (1959). The science of "muddling through". *Public Administration Review*, 19(2), 79–88.
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *Journal of Intelligent Systems*, 16(2), 72–79.
- Malzahn, N., Urspruch, T., Zeini, S., & Hoppe, H. U. (2007). Dynamisierung von personal- und kompetenzentwicklung. In K. Shire, M. Borchert, & H. U. Hoppe (Eds.), *Jenseits von Virtualität* (pp. 229–255). Lohmar: EUL Verlag.

- Medin, D., Goldstone, R., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100(2), 254–278.
- Online Vermarkter Kreis. (2009). *Overview of figures and trends*. Bundesverband Digitale Wirtschaft. Retrieved from <http://www.bvdw.org/mybvdw/media/download/ovk-online-report-200901-english.pdf?file=405>. Berlin: BVDW e.V.
- Piazza, F., & Strohmeier, S. (2011). Domain-driven data mining in human resource management: A review. *Proceedings of the 11th IEEE International Conference on Data Mining Workshops* (pp. 458–465). IESS.
- Porter, M. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137.
- Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, 19(1), 17–30.
- Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference for Artificial Intelligence*.
- Rodrigues, S., Oliveira, J., & de Souza, J. M. (2004). Competence mining for virtual community creation. *International Journal of Web Based Communities*, 1(1), 90–102.
- Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613–620.
- Schröder, S. (2007). *Implementierung und evaluation eines modells zur dynamisierung von personal- und kompetenzentwicklung in der IT-branche*. Duisburg: University Duisburg-Essen.
- Shire, K., Borchert, M., & Hoppe, H. U. (Eds.). (2007). *Jenseits von virtualität - Arbeiten und lernen in projektnetzwerken*. Lohmar: EUL Verlag.
- Strauss, A. L. (1993). *Continual permutations of action*. New York: Walter de Gruyter.
- Sure, Y., Maedche, A., & Staab, S. (2000). Leveraging corporate skill knowledge - From proper to ontoproper. *Proceedings of the third International Conference on practical aspects of Knowledge Management*.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 4.
- Vygotski, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Weinbrenner, S., Giemza, A., & Hoppe, H. (2007). Engineering heterogeneous distributed learning environments using Tuple Spaces as an architectural platform. *Proceedings of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT)*. Los Alamitos, USA: IEEE Computer Society.
- Weinert, F. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 45–66). Göttingen: Hogrefe & Huber.
- Witten, I., & Frank, E. (2005). *Data mining - Practical machine learning tools and techniques*. Elsevier Inc.
- Wittgenstein, L. (1953). *Philosophical investigations*. Blackwell Publishing.
- Woodruffe, C. (1993). What is meant by a Competency. *Leadership & Organization Development Journal*, 14(1), 29–36.
- Zelewski, S. (2005). Kooperatives wissensmanagement in engineering-netzwerken (KOWIEN) - Industrieerprobte lösungen und werkzeuge für produktentwicklung, engineering and kompetenzmanagement. *Proceedings zum Abschlussworkshop der Verbundprojekte GINA, KOEFFIZIENT und KOWIEN* (pp. 21–83). Braunschweig, Germany.
- Zhu, J., Goncalves, A. L., Uren, V. S., Motta, E., & Pacheco, R. (2005). Mining web data for competency management. *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 94–110).

- Ziebarth, S., Malzahn, N., & Hoppe, H. U. (2009). Using data mining techniques to support the creation of competence ontologies. *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED)* (pp. 223–230). Brighton, England.
- Ziebarth, S., Malzahn, N., Daems, O., & Hoppe, H. U. (2010). Creation and matching of competence profiles in the German digital economy using semantic methods. In J. Hafkesbrink, H. Hoppe, & J. Schlichter (Eds.), *Competence Management for Open Innovation. Tools and IT support to unlock the innovation potential beyond company boundaries* (pp. 69–92). EUL Verlag.