Extending Formal Concept Analysis by Fuzzy Bags

Paolo Ceravolo
Dipartimento di
Tecnologie dell'Informazione
Università degli studi
di Milano
ceravolo@dti.unimi.it

Ernesto Damiani
Dipartimento di
Tecnologie dell'Informazione
Università degli studi
di Milano
damiani@dti.unimi.it

Marco Viviani
Dipartimento di
Tecnologie dell'Informazione
Università degli studi
di Milano
viviani@dti.unimi.it

Abstract

This paper proposes an extension of Formal Concept Analysis (FCA) techniques in order to deal with representation of concepts in terms of fuzzy bags. Our extension was tested in a preliminary experimentation with encouraging results. In particular, the proposed technique was adopted as a tool for automatic quick-and-dirty ontology construction.

Keywords: Ontology Construction, Formal Concept Analysis, Fuzzy Bags.

1 Introduction

The Semantic Web vision relies on ontologies to structure its underlying data for supporting basic and advanced machine understanding. Automatic learning of ontology has been identified since long as a key issue for integrating diverse sources of unstructured, semi-structured, and fully structured information. Also, bottom-up learning of ontologies from text documents is a fundamental support for modeling the domain of interest of an organization [13]. Tools like Text-To-Onto [14] are now available that attempt to learn structured ontologies from free text, dictionaries, or legacy domain model. Text-to-ontology learning methodologies rely on basic techniques for text document classification, and usually start by indexing text documents via vectors of (normalized) keyword occurrences. Thus, documents can be classified using a standard *vector space model* where every document d, is considered to be a vector $\boldsymbol{d}$ in the term space, i.e. the overall vocabulary. In this scenario, each document is represented by the (TF) vector $d_{ft} = (f_{t_1}, f_{t_2}, ..., f_{t_n})$ where f_{t_i} is the frequency of the i^{th} term in the document. In order to account for documents of different lengths, each document vector is normalized so that it is of unit length. This encoding is used in order to group similar documents by means of distance measures, such as the well-known cosine distance. These groups of documents, loosely called clusters, represent typical classes of the domain. In ontology learning, however, a further step is required, because domain classes must be ordered into a hierarchy. Some approaches rely on preliminary linguistic processing, where terms are analyzed, aggregated into concepts, and progressively organized according to taxonomical relations and rules ([1], [8], [7], [12]). Other extraction methodologies apply *Formal Concept Analysis* (FCA), a time-honored technique used to build hierarchies of common subset of attributes from a set of data items [20]. Namely, the concept hierarchy is obtained applying an algorithm (several versions exist, see [17], [10], [19]) to derive a lattice of shared terms within the document set. A major problem of FCA is related to the identification of noise, i.e. irrelevant information. Even the analysis of a small document set require to deal with a large number of terms; very often, most of them are not interesting for the analysis.

In this paper we propose a method for extending FCA to represent document vectors in terms of fuzzy bags. We claim that our extension can ameliorate the quality of the hierarchy produced, because additional information related to relevance and cardinality of individual data items' attributes can be taken into account that would be lost by traditional methods. This additional information can be used in order to partition concepts of the hierarchy in sub-concepts differentiated according to the relevance and the cardinality of attributes. According to our proposal, a term vector is described as a *fuzzy bag* [22], [23]. Fuzzy bags are a straightforward extension of fuzzy sets, where each element can have multiple instances. In a fuzzy set each element is associated to a membership value. In our setting, multiple membership values can be associated to a single term in order to expressing the relevance of a term in the vector. This way, individual terms can be represented as a single entry, while still taking into account cardinality differences; also, irrelevant terms can be easily discarded. The paper is organized as follows: Section 2, analyzes the research works related to our proposal; Section 3 gives an introduction to the problem of representing vectors of terms in terms of fuzzy bags; Section 4 provides a formal definition of our extension of FCA; Section 5 provides an example of ontology construction by means of our method; Section 6 describes the results achieved and further research lines.

2 Related work

Different approaches have been proposed to integrate FCA and fuzzy logic. In [2] and in [11] the set of truth values has the structure of a complete residuated lattice, where Galois connections are expressed in terms of fuzzy binary relations. This approach provides a very expressive representation of the correlation between documents and terms but it is not very compliant to the common output of knowledge extraction applications. In [15] the notion of a *L-Fuzzy context* is proposed, where linguistic variables are used in order to represent uncertainty in the term-document relationship. However, linguistic variables can be defined only on the basis of human interpretation, and this approach turns out not to be feasible when dealing with very large document sets. In [16] a technique is proposed called *Fuzzy Formal Concept Analysis* (FFCA), in which term relevance is represented by a membership value in the range $[0, 1]$. This way, documents are not described as term vectors but as fuzzy sets, where each term is associated to a membership value expressing its relevance w.r.t. the document set. In this paper we extend this notion, representing documents as fuzzy bags and extending the operations on the document set according to the operations on fuzzy bags.

3 Representing term vectors as fuzzy bags

Fuzzy bags can be used for extending the traditional representation of documents as vectors of terms. Indeed, a conventional term vector has the limitation of not taking into account the cardinality of term relevances. For example, keywords extracted from a document can be used to infer the topics the document deals with, but the association between keywords and topics cannot be treated equally in all cases, because keywords may have a different degree of significance for different topics. For this reason, a relevance value can be modeled by means of a fuzzy membership function. For example a document associated to a set of keywords composed by "clustering", "ontology", and "fuzzy set" can be related to the topic *ontology learning* with a membership degree equal to 0.8, but with the topic *fuzzy set theory* with a degree equal to 0.3. Other method can be used in order to assign a relevance value to a term. The general idea is to evaluate the informativeness of a term in a document. The conventional approach is to compute a ratio between the frequency of the term in the analyzed document and in the whole domain vocabulary. But other approaches take into account the context in which a term is inserted; and in this case different relevance values can be associated to a single term.

Fuzzy bags are an interesting tool in order to model multiple relevance values, because they represent the distribution of the cardinality of an element according to different membership values. In [6] a method for representing semi structured-data (e.g. XML documents) in terms of fuzzy bags is proposed. The rationale of this approach is encoding informativeness of XML elements according to their structural position. A preliminary method for evaluating the informativeness of an element is to divide a membership degree initially equal to 1 by the nesting level of the element. Fig. 1 shows a cluster of XML documents encoded according to this method:

$A = \{1/R,\ 0.5/a,\ 0.5/b\}$,
$B = \{1/R,\ 0.5/a,\ 0.5/b,\ 0.3/b\}$,
$C = \{1/R,\ 0.5/a,\ 0.5/b,\ 0.3/c,\ 0.3/c\}$,
$D = \{1/R,\ 0.5/a,\ 0.5/b,\ 0.3/c,\ 0.3/c\}$,
$E = \{1/R, 0.5/a,\ 0.5/b,\ 0.3/c,\ 0.3/a\}$.

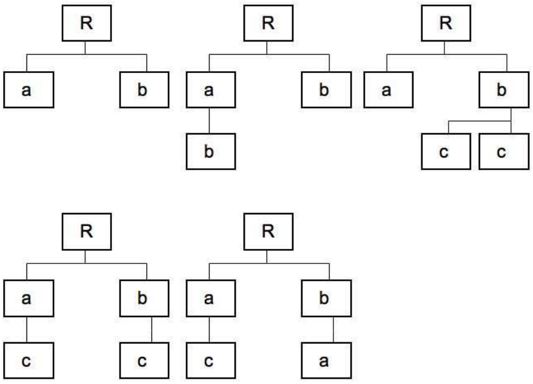

Figure 1: Representation of XML documents in a cluster

Using the *FGcount* notion of cardinality proposed by Zadeh in [25], the cardinality of a fuzzy bag can be represented listing a *membership-value/cardinality* pair obtained applying on the fuzzy bag an α-cut operator for each membership-value from 0 to 1. According to this definition the XML documents in Fig 1 can be encoded as follows:

$A = \{\{0.5/1\} * a,\ \{0.5/1\} * b\}$,
$B = \{\{0.5/1\} * a,\ \{0.5/1,\ 0.3/2\} * b\}$,
$C = \{\{0.5/1\} * a,\ \{0.5/1\} * b,\ \{0.3/2\} * c\}$,
$D = \{\{0.5/1\} * a,\ \{0.5/1\} * b,\ \{0.3/2\} * c\}$,
$E = \{\{0.5/1,\ 0.3/2\} * a,\ \{0.5/1\} * b,\ \{0.3/1\} * c\}$.

This construction is adopted in [5] for defining fuzzy bags in a way fully compatible with fuzzy sets. This construction proposes the notion of *gradual integer*, because it is able to represent the cardinality of a fuzzy bag according to the distribution of different membership degrees. In [4] a complete discussion on gradual integers and the corresponding arithmetic operations is provided, on the basis of these arithmetic operations a set of set-theoretical operations on fuzzy bags is proposed. Relying on these works we will be able to manage the representations of terms vectors in terms of fuzzy bags.

4 Fuzzy Formal Concept Analysis

Standard FCA techniques build a *concept lattice* by organizing a binary relation over a pair of sets $\mathcal{D}$ and $\mathcal{A}$, where $\mathcal{D}$ represents the data items (called *documents* in he FCA terminology) and $\mathcal{A}$ the attributes [20].

Definition 1. A formal context is a triple $\mathcal{K} = (\mathcal{D}, \mathcal{A}, \mathcal{R})$ where $\mathcal{D}$ and $\mathcal{A}$ are sets and $\mathcal{R}$ is a binary relation $\mathcal{R} \subseteq \mathcal{D} \times \mathcal{A}$.

Table 1 shows a description of a formal context. The numbers in the left column indicate documents belonging to the document set $\mathcal{D}$ while the letters in the first row indicate the attributes $\mathcal{A}$. A true boolean value in a cell belonging to a row $d \in \mathcal{D}$ and a column $a \in \mathcal{A}$ means that the document d contains the attribute a.

The table can also be summarized by two functions: f and g, defined as follows.

Definition 2. The function f maps a set of documents into set of common attributes, whereas g is the dual for the attribute sets.

If X and Y are respectively sets of documents and attributes:

Table 1: An sample formal context visualized in tabular form.

0	a	b	c	d	e
1	1	1	0	0	1
2	1	0	1	1	1
3	0	1	1	1	1
4	1	0	0	1	0
5	0	0	1	1	1
6	1	1	1	0	1
7	1	1	0	1	0

$$f(X) = \{a \in A | \forall d \in X, d\mathrm{R}a\} \tag{1}$$

$$g(Y) = \{d \in D | \forall a \in Y, d\mathrm{R}a\} \tag{2}$$

For example in the context of Table 1 $f(35) = cde$ and $g(ab) = 167$. A couple (X, Y), of mutually corresponding closed subsets is called a *formal concept*.

Definition 3. A formal concept ia a couple (X, Y) where $X = g(Y)$ and $Y = f(X)$. X is called the *extent* and Y the *intent* of the concept.

For example $(167, ab)$ is a formal concept but $(167, a)$ or $(23, cde)$ are not. The set of coupled documents and attributes are not the ones that can be obtained by applying f or g respectively.
These definitions allows us to define the lattice formed organizing in a partial order the formal concepts belonging to a formal context.

$$\bigvee_{i=1}^{k}(X_i, Y_i) = (\bigcup_{i=1}^{k} f(X_i), \bigcap_{i=1}^{k} Y_i), \tag{3}$$

$$\bigwedge_{i=1}^{k}(X_i, Y_i) = (\bigcap_{i=1}^{k} X_i, \bigcup_{i=1}^{k} g(Y_i)). \tag{4}$$

Note that, for instance, concept $(235, cde)$ has the super-concept $(2356, a)$ and the subconcept $(2, acde)$.

Now, let us present in some detail our extension of FCA methodologies by interpreting documents as fuzzy bags.

Definition 4. A formal context extended to deal with fuzzy bags is a triple $\mathcal{K}_{ex} = (\mathcal{D}, \mathcal{A}, \mathcal{R}_{ex})$ where $\mathcal{D}$ and $\mathcal{A}$ are sets and $\mathcal{R}_{ex}$ is a fuzzy bag on domain D × $\mathcal{A}$. Where each relation (d, a) $\in \mathcal{R}_{ex}$ is a gradual integer $\Omega_d(a)$.

In FCA the codomain of the function f, i.e. the set of attributes common to a given set of documents X, is equivalent to the intersection of the documents in X, and identifies a set of common attributes Y. Hence, in our extension, the function f_{ex} has to return the intersection among the fuzzy bags associated to the documents in X, and the fuzzy bag obtained by this intersection must be included by all the fuzzy bags of X, as formally described in equation 5.

$$\Omega_{f_{ex}(X)} \subset \Omega_i, \forall i \in X \tag{5}$$

According to [5], equation 6 gives us the definition of intersection among two gradual integers.

$$\Omega_{A \cap B}(x) = min(\Omega_A(x), \Omega_B(x)) \tag{6}$$

Because minimum is an associative operation, we can apply it to a set of attributes. In other words, in our extension the function f_{ex} maps a set of documents to a fuzzy bag generated by the intersection of gradual integers associated to their common attributes, For each attribute $a \in Y$ the gradual integer of a is computed according to the equation 7.

$$\Omega_{f_{ex}(X)}(a) = \min_{i \in X} \Omega_i(a) \tag{7}$$

Similarly, the function g_{ex} maps a fuzzy bag $\Omega_{g_{ex}(Y)}$ generated by the minimum of the gradual integers, associated to a set of attributes Y, to all the documents described by a fuzzy bag including $\Omega_{g_{ex}(Y)}$.

$$\Omega_{g_{ex}(Y)}(a) = \min_{i \in Y} \Omega_i(a) \qquad (8)$$

5 Building the hierarchy

Our extension of FCA techniques was experimented with a dataset selected from the *XML Data Repository* of the University of Washington [26][1]. This repository hosts entries belonging to different digital archives of research publications, and therefore needs little pre-processing in terms of stop-word filtering besides disregarding connective particles such as prepositions and articles. Also, it is a typical example of heterogeneous data in semi-structured format. As alternative examples, we might consider the publications entries from the **DBLP** and from the **Sigmod** databases.

We start by encoding these documents as fuzzy bags [6]; then, a clustering of the document set is executed. Each cluster is identified by a representative, called *cluster-head*, usually equivalent to the intersection of all documents belonging to the cluster. Cluster-heads are used to define the initial context of our FCA analysis. Being fuzzy bags, cluster-heads are composed of *term/ gradual integer* pairs, as discussed in our previous example. Table 2 shows the context formed by our cluster-heads. For the sake of conciseness, in the sequel fuzzy bags' elements (the terms of the context's vocabulary) are denoted by the initial letters of the element name; also, some attributes are omitted.

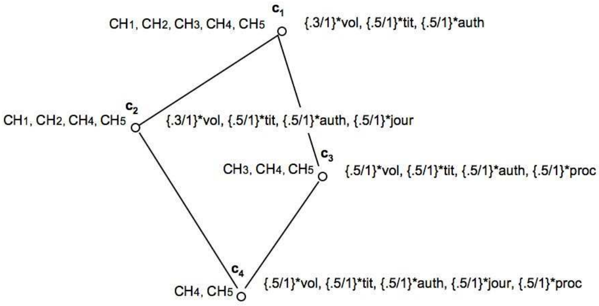
c1
CH1, CH2, CH3, CH4, CH5
{.3/1}*vol, {.5/1}*tit, {.5/1}*auth
c2
CH1, CH2, CH4, CH5
{.3/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour
c3
CH3, CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*proc
c4
CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour, {.5/1}*proc

Figure 2: The lattice provides a first hierarchy of candidate classes.

Table 2: The context of our cluster-heads.

	vol.	tit.	auth.	jour.	proc.
CH_1	0.3/4	0.8/4	0.8/4	0.5/4	0
CH_2	0.5/19	0.8/19	0.8/19	0.5/19	0
CH_3	0.5/1	1/1, 0.5/5	0.5/5	0	1/1
CH_4	0.5/1	0.5/1	0.5/1	0.5/1	0.5/1
CH_5	0.5/1	1/1, 0.5/12	0.5/23	1/1	0.5/6

Applying f_{ex} and g_{ex} we obtain the hierarchy in Fig. 2. Note that hierarchy isolated $c1$ as the subset of the attributes common to all the documents of our context. In $c2$ we have documents published in proceedings, in $c3$ documents published in a journal, and in $c4$ documents published both in proceedings and journals. Note that this hierarchy could have been obtained by standard FCA techniques; but in our extension gradual integers encode information about relevance and cardinality of terms. Now by this information we can retrieve additional formal concepts, further partitioning the ones obtained in the first hierarchy. For instance, we can consider cardinality as a modifier of the extent of a concept and we can require that documents having strong cardinality differences belong to different formal concepts. This can be done by comparing the intent of each document belonging to a formal concept to the intent of the concept (that is, the fuzzy bag corresponding to the formal concept). All documents having strong relative differences in cardinality will form the intent of a new concept, while the corresponding extent is obtained computing the intersection among their attributes according to equation 7. In order to do that, we must be able to provide a semantics to the predicate "having relevant attributes with strong differences in cardinality". In ordinary arithmetics, two numbers are equal if dividing one to the other we get a result equal to 1. The same principle

[1] Note that for our present purposes, XML tag names have no special status w.r.t. content terms.

can be applied to gradual integers but we need to approximate them in an exact reperesentation. According to the notion of cardinality exposed in [21], the cardinality of a fuzzy set can be interpreted as the exact number summing all the cardinality of a set weighted by their membership values. Formally we have:

$$|A| = \sum_{x_i \in A} f(\mu(X_i)) \tag{9}$$

Where f is an increasing function in $[0,1]$, where $f(0) = 0$ and $f(1) = 1$. For instance, $f(x) = x^2$. This way the relative weight of higher membership values is increased.

Now we can compute the relative difference between the exact numbers approximating the gradual integers related to two attributes of a document. This result is compared to the number 1 and we obtain a measure that can be used to define the semantics of the predicate “having relevant attributes with differences in cardinality”. This can be dome by the complement of the relative difference. Formally we have:

$$1 - \frac{|\Omega_i(a)|}{|\Omega_j(a)|} \tag{10}$$

In our example, the formal concept $c2$ has an intent composed by the documents CH_1 and CH_2. Comparing the extent of CH_2 to the extent of $c2$ we have to calculate the difference between each single attribute. For instance we get $|\{0.3/1 * vol\}|/|\{0.5/19 * vol\}| = 0.01$. The complement of this division is $1 - 0.01 = 0.99$. Similar results are obtained for the other attributes. Now, computing an average among the results of each attribute, we obtain a distance equal to 0.98. This value allow us to said CH_2 “has strong differences in cardinality” with $c2$ with a degree of 0.98. This motivate us to generate a new formal concept. A similar analysis can be done with the document CH_5 as shown in Fig. 3.

A further step consists in pruning the hierarchy to eliminate irrelevant attributes. This can be done setting a *threshold* on the membership values associated to the attributes of the fuzzy bags forming the context. For instance, if the confidence threshold T is equal to 0.4 we impose to the formal context the condition expressed in equation 11. The result is the removal of the attribute `volume` from some formal concepts of the hierarchy.

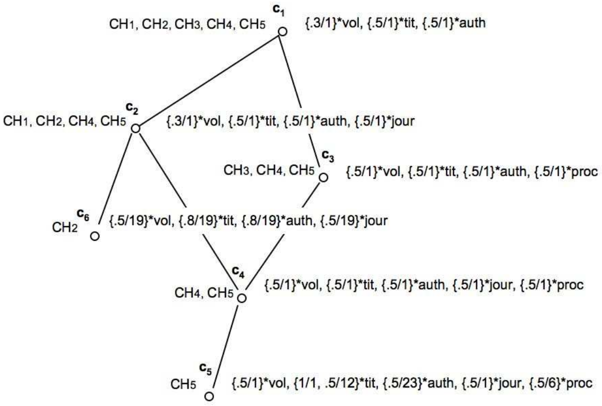
c1
CH1, CH2, CH3, CH4, CH5
{.3/1}*vol, {.5/1}*tit, {.5/1}*auth
c2
CH1, CH2, CH4, CH5
{.3/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour
c3
CH3, CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*proc
c6
CH2
{.5/19}*vol, {.8/19}*tit, {.8/19}*auth, {.5/19}*jour
c4
CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour, {.5/1}*proc
c5
CH5
{.5/1}*vol, {1/1, .5/12}*tit, {.5/23}*auth, {.5/1}*jour, {.5/6}*proc

Figure 3: A hierarchy where cardinality can generate new formal concepts.

$$\mu(a, d) \geq T \tag{11}$$

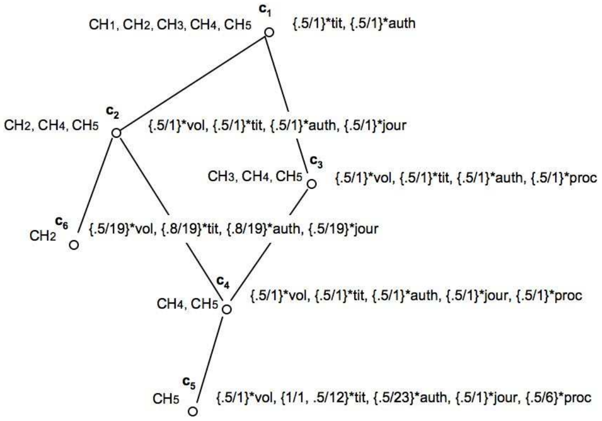
c1
CH1, CH2, CH3, CH4, CH5
{.5/1}*tit, {.5/1}*auth
c2
CH2, CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour
c3
CH3, CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*proc
c6
CH2
{.5/19}*vol, {.8/19}*tit, {.8/19}*auth, {.5/19}*jour
c4
CH4, CH5
{.5/1}*vol, {.5/1}*tit, {.5/1}*auth, {.5/1}*jour, {.5/1}*proc
c5
CH5
{.5/1}*vol, {1/1, .5/12}*tit, {.5/23}*auth, {.5/1}*jour, {.5/6}*proc

Figure 4: Pruning the hierarchy by means of a confidence threshold T.

The hierarchy obtained was used as an initial representation of the domain of scientific publications. Ontology engineers has to manually develop this tentative representation in order to obtain an ontology fitting a standard format.

6 Conclusions

In this paper we proposed an extension of FCA techniques exploiting a fuzzy representation of documents in terms of fuzzy bags. The adoption of fuzzy bags is motivated by the opportunity to encode information related to relevance and attributes cardinality in a simple format. We proposed a formal redefinition of FCA according to our extension. Then, we introduced a preliminary experimentation of our technique in the field of ontology construction. In this filed the adoption of our extension is motivated by the capability to describe tentative classes according to additional feature, such as cardinality or relevance of attributes. Our preliminary results are encouraging and point to a promising line of research investigating the notion of order relation among gradual integers to derive new operations to be applied on the formal context.

Acknowledgements

This work was partly funded by the Italian Ministry of Research Fund for Basic Research (FIRB) under projects RBAU01CLNB_001 "Knowledge Management for the Web Infrastructure" (KIWI).

References

[1] K. Ahmad and A. E. Davies, Weirdness in Special-language Text: Welsh Radioactive Chemicals Texts as an Exemplar. Internationales Institut fr Terminologieforschung Journal: 22–52, 1994.

[2] R. Belohlavek, Fuzzy Galois Connections. Math Logic Quarterly, 45(4): 497–504, 1999.

[3] R. Belohlavek and V. Sklenar, Formal Concept Analysis Constrained by Attribute-Dependency Formulas, Lecture Notes in Computer Science, 3403: 176–191, 2005.

[4] D. Rocacher and P. Bosc, The set of fuzzy rational numbers and flexible querying. Fuzzy Sets and Systems, 2005.

[5] D. Rocacher and P. Bosc, About Zf, the Set of Fuzzy Relative Integers, and the Definition of Fuzzy Bags on Zf. IFSA, Springer-Verlag (2715): 95–102, 2003.

[6] P. Ceravolo, M. C. Nocerino and M. Viviani, Knowledge extraction from semi-structured data based on fuzzy techniques. Knowledge-Based and Emergent Technologies Relied Intelligent Information and Engineering Systems: 1257–1263, 2004.

[7] D. Faure and T. Poibeau, First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX. In Proceedings of the ECAI Workshop on Ontology Learning, 2000.

[8] P. Gamallo, M. Gonzalez, A. Augustinin, G. Lopes and V. S. de Lima, Mapping Syntactic Dependencies onto Semantic Relations. In Proceedings of 15th European Conference on Artificial Intelligence Workshop on Machine Learning and Natural Language Processing for Ontology Engineering, 2002.

[9] G. Georgescu and A. Popescu, Non-commutative fuzzy Galois connections. Soft Computing, Springer-Verlag, 7: 458–467, 2003.

[10] G. Jiang et al, Concept-oriented view generation for clinical data using formal concept analysis. Proceedings of JCMI2004, Nagoya, Japan, 2004.

[11] A. Jaoua, F. Alvi and S. Elloumi, Galois Connection in Fuzzy Binary Relations, Applications for Discovering Association Rules, SB Yahia, RelMiCS, 2000.

[12] U. Hahn and K. G. Marko, Ontology and lexicon evolution by text understanding. In Proceedings of the ECAI Workshop on Machine Learning and Natural Language Processing for Ontology Engineering, 2002.

[13] J.U. Kietz, A. Maedche and R. Volz, A method for semi-automatic ontology ac-

quisition from a corporate intranet, Proceedings of the EKAW Workshop on Ontologies and Text, 2000.

[14] A. Maedche and S. Staab, Ontology Learning for the Semantic Web, IEEE Intelligent Systems 16 (2) : 72 – 79, 2001.

[15] S. Pollandt, Fuzzy-Begriffe: Formale Begriffsanalyse unscharfer Daten, Springer. Verlag, Berlin, 1996.

[16] T. T. Quan, S. C. Hui and T. H. Cao: A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data. CLA, 2004.

[17] G. Stumme and A. Maedche, FCA-merge: bottom-up merging of ontologies. In Proc. 17th IJCAI, Seattle: 225–230, 2001.

[18] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier and L. Lakhan, Computing iceberg concept lattice with Titanic. Journal on Knowledge and Data Engineering, 42(2): 189–222, 2002.

[19] G. Toa, Using Formal Concept Analysis (FCA) for Ontology Structuring and Building. PhD thesis, 1992.

[20] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, Ordered Sets: 445–470, 1982.

[21] M. Wygralak, An axiomatic approach to scalar cardinlities of fuzzy sets, Fuzzy Sets System, 110, 175–179, 2000.

[22] R. Yager, On the theory of bags, Internat. J. Gen. Sys, 13(47): 23–27, 1986.

[23] R. Yager, Cardinality of fuzzy sets via bags, Math. Modelling, 9(6): 441–446, 1987.

[24] R. Yager and D. Filev, On the issue of defuzzification and selection based on a fuzzy set, Fuzzy Sets System, 55(3) 255–271, 1993.

[25] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Computer Mathematics with Applications, 9, p. 149-183, 1983.

[26] XML Data Repository, University of Washington, http://www.cs.washington.edu/research/xmldatasets, 2005.