



HAL
open science

Conflict Ontology Enrichment Based on Triggers

Chahnaz Zakaria, Olivier Curé, Kamel Smaïli

► **To cite this version:**

Chahnaz Zakaria, Olivier Curé, Kamel Smaïli. Conflict Ontology Enrichment Based on Triggers. The 2nd International workshop on Ontologies and Information Systems for the Semantic Web, ACM 17th Conference on Information and Knowledge Management, Oct 2008, Napa Valley, California, United States. inria-00315300

HAL Id: inria-00315300

<https://inria.hal.science/inria-00315300v1>

Submitted on 28 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conflict Ontology Enrichment Based on Triggers

Chahnez Zakaria

Laboratory S3IS
Université Paris Est
5, bd Descartes Champs sur Marne
77454 Marne la Vallée France
czakaria@etudiant.univ-mlv.fr

Olivier Curé

Institut Gaspard Monge
Université Paris Est
5, bd Descartes Champs sur Marne
77454 Marne la Vallée France
ocure@univ-mlv.fr

Kamel Smaïli

Laboratory Loria
Campus Scientifique
BP 239 54506
Vandoeuvre Lès-Nancy
smaïli@loria.fr

Abstract

In this paper, we propose an ontology-based approach that enables to detect the emergence of relational conflicts between persons that cooperate on computer supported projects. In order to detect these conflicts, we analyze, using this ontology, the e-mails exchanged between these people.

Our method aims to inform project team leaders of such situation hence to help them in preventing serious disagreement between involved employees.

The approach we present builds a domain ontology of relational conflicts in two phases. First we conceptualize the domain by hand, then we enrich the ontology by using the trigger model that enables to find out terms in corpora which correspond to different conflicts.

Categories and Subject Descriptors

H. Information Systems

General Terms

Algorithms, Management.

Keywords

Ontology, Corpus, Triggers, Mutual Information.

1. INTRODUCTION

1.1 Application Area

The nature and organization of the companies' activities have changed considerably since the integration of Information and Communication Technologies (ICT) and more exactly with their evolution to distributed architecture. As a result, virtual organizations have emerged, with virtual teams composed of employees distributed geographically around the world. These team members must then collaborate and communicate via Computer Supported Cooperative Work (CSCW) tools, to thwart the problem of distance.

The integration of ICT has also changed the concept of enterprise management, however, managing employees emotions remains

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ONISW'08, October 30, 2008, Napa Valley, California, USA.
Copyright 2008 ACM 978-1-60558-255-9/08/10...\$5.00.

one of the concerns of decision-makers. It is well known that conflicts coming out in such contexts can heavily influence the performance and efficiency of members of these companies. Studying employees' behavior on a productivity context owes its origin to Hawthorne's experiences inspired by Elton Mayo [1]. Basically these experiences were designed to study the influence of physical conditions on employees' productivity.

These experiences also included the study of communication barriers, the cooperation's phenomena and conflict between employees, what allowed to deduce that good horizontal and/or vertical relationship had a major influence on the global satisfaction of employees and on their personal productivity.

1.2 Problem and Proposal

The constitution of virtual team has increased the difficulty of tracking the cooperation between members and managing their emotions. Because of geographical distance, the team leader found itself confronted with problems of understanding the employees behavior: how can he remotely detect the possible situations of conflicts between them? He has mainly to become aware of the conflict before it came up, otherwise the clash will influence their communication and their cooperation?

The geographically distributed enterprises have adopted the use of CSCW tools, to overcome the problems of distance, as *shared publishers*, *shared whiteboards*, *audioconference*, *videoconference*, *forums*, *chat*, *electronic mail*, etc.

These tools may help the leader's team to be informed about the cohesion of the group, especially the e-mail, it allows to generate the most important textual corpora, because it is the most used [2].

In order to better understand and specify the needs of users' electronic mails, we adopted the realistic approach of Kyng (stemming from Scandinavian school) [4]. This approach puts building tools in the center of close collaboration between developers and users.

We collaborated with an Algerian university¹ with the intention of specifying the problems of our research. Our jobs concerned the management of a remote training.

A multidisciplinary team has been set up to prepare this remote training. For that, they mediatize (design multimedia versions) educational contents. The members of this team communicate through e-mails.

We studied the corpus generated by the team of educational content mediatization, and we identified several messages that are subject to creating conflict situations. This situation is due to time

¹ Université de la Formation Continue

constraints which submit the members of the team to a high stress. Figure 1 gives an example about an e-mail which creates a conflict, the primary text is in French. It is part of an e-mail sent by a coordinator (a teacher who follows the work of mediatisors) to a mediatisor (a teacher or a computer engineer who have good knowledge on how to create content for e-learning). The terms on box in figure 1 shows explicitly that necessary a problem is rising between two people. In this case, the team leader must mediate to calm down the two protagonists.

The team members put their head-chief in copy for all e-mails exchanged between them, allowing him to monitor their cooperation. Obviously, this will generate a huge amount of e-mails and the head could not follow all the exchanges between all people.

It should be very attractive to inform automatically the leader only when a problem starts rising. Then, the leader can intervene in before the situation becomes irreversible. Consequently, the system we propose has to be able to understand the behavior of an employee through e-mails and to find out all tracks which lead to conflict with his colleagues.

To detect conflicts, we propose to classify e-mails, according to the existence or missing of any conflict indicator. Then, we determine the nature of clash through a conflict domain ontology by assigning a value which indicates the importance of the clash. Indeed, it is not necessary to alert the team leader for any misunderstanding between people, but we have to send him a list of e-mails labeled by the degree of the clash. This will allow him to decide or not to intervene.

In this paper, we present the approach we have adopted to build the conflict ontology.

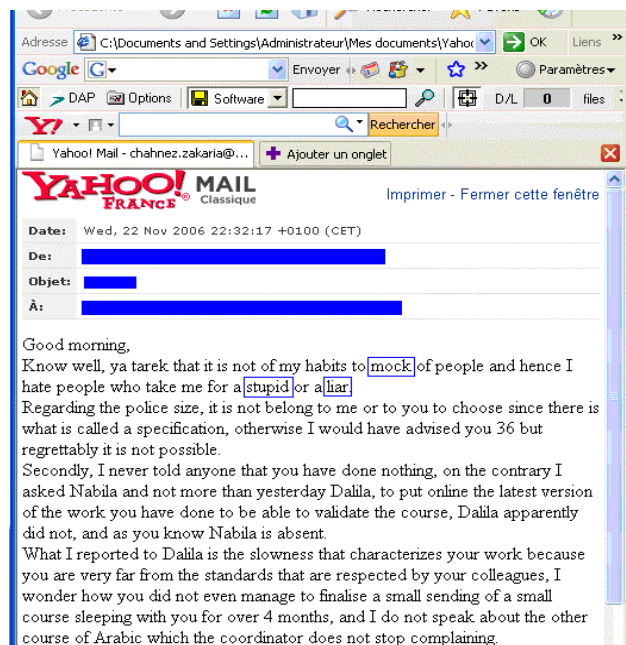


Figure 1. An e-mail which creates a conflict between a coordinator and a mediatisor

2. RELATED WORK

Several approaches are proposed for building ontologies from corpora. They can be grouped into three categories: structural approaches based on the use of formal grammar; non-structural approaches, such as statistical approaches which are more and more used due to the availability of corpora; and mixed approaches involving statistical analysis and structural methods as those used on *Terminae* [9], *Xtract* [11], *KAON* [10], etc.

For the acquisition of terms, structural approaches are based on morphological-syntactic labeling or complete parsing of the corpus. They can be divided into two approaches: those which select candidates terms on the basis of a canonical syntactic structure of a term as *acabit* [17] and *OntoLearn* [12] tools; and those which identify marks of ruptures terms, i.e. configurations that can not be part of a term, to give the borders (strong punctuation and pronoun for example), as *Lexter* [15] and *Syntex* [16] tools.

Indeed structural approaches often require grammar and sometimes lexicons or electronic dictionaries of the used language, unlike the statistical approaches that are blind to the specificity of linguistic data, they are treated as any other type of data [7]. However statistical approaches must use enough corpora in order to have reliable measures and find out interesting relationships between terms [8].

The acquisition of terms based on statistical approach exist since several decades: Enguehard and Pantera (1995) [18], Dias (2002) [2], etc. This work is based on the idea that words of the same area tend to often occur together. Similarity measures are used to identify recurrent associations of terms. The correlated terms recurrences are extracted by using different kind of measures [9] [10]: Mutual Information, Dice coefficient, etc.

3. BUILDING ONTOLOGY

3.1 Construction of a first conflict ontology draft

The main characteristics of Knowledge-Based Systems are storage and consultation of knowledge, automatic reasoning, modification of knowledge, sharing knowledge between computer systems. Generally, it is not any more a question of handling blind knowledge by the machine, but of allowing a dialog between the system and the human user. The system thus has to access not only to the terms used by the human being, but also to meaning associated to the various terms, otherwise no effective dialog is possible. So ontology technology was born as a response to the need for representation of knowledge in information systems. For instance T. Berners-Lee [3] considers the ontology as a way enabling Web pages to integrate a representation of knowledge they contain, and semantic links between them.

The most cited and famous definition of ontology in the scientific literature is the one proposed by Gruber in 1993: "ontology is an explicit specification of a conceptualization" [6]. In other words, an ontology is designed to specify concepts and relations, i.e. to make them understandable and usable by several agents (human or software).

One of the fundamental questions when we build an ontology is how to do it? Automatically, semi-automatically or by hand?

For our ontology of conflicts the choice was focused on semi-automatic way. A first step consists in building a rough version by using human expertise which leads to conceptualize the area of conflict. And in a second step we enrich it automatically from a textual corpus.

In our knowledge, there is no taxonomy of conflicts, that is why we decided to focus our work on emotions. In fact, conflicts could be detected through the expression of their emotions. Here, we are interested just by negative emotions which are at the origin of disagreement situations. For that, we based our work on the classification of Antonio R Damasio [13] and Michelle Larivey [14]. We used their vocabularies, but we changed the separation criteria of emotions. Concepts are classified according to the degree of conflict and to the type of emotion (personal or social).

We organize the first level of our ontology on two categories:

- The first category represents emotions that can produce substantial conflict as disgust and hatred.
- The second one leads to anticipate some indirect conflicts due to indifference, misunderstanding, etc.

The second level, we make the difference between personal and social emotions. Actually, it is very difficult to distinguish between them. For instance, the sadness emotion may be social when this feeling is due to the behavior of another a colleague or friend and may be personal when the person did not succeed to reach an objective. Some emotions can be categorized easily in the social one, that is the case of jealousy which is by definition a social concept. We have presented in figure 2 other concepts of social emotions.

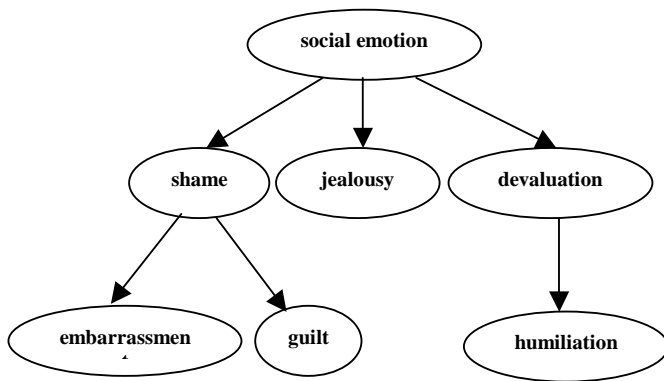


Figure 2. An extract of our conflict ontology

We used the e-mail corpus presented in the following sections, to find out other terms that match perfectly with the real conflict situation. To enrich our ontology, we seek in the e-mail corpus semantically related words using the triggers that are introduced in the next section.

3.2 Triggers to Enrich Ontology

Development of statistical language models is historically related to the construction of the first significant linguistic corpora. Indeed statistical methods were known as early as the 1950 but they have been really used in the Natural Language Processing when the first significant linguistic corpora has been available [7] [20].

For these models, a corpus represents a raw material, it is used to learn a maximum of linguistic events (n-grams, part of speech, etc) [5] [8]. In other words statistical processing of corpora allows to get knowledge by studying recurrent phenomenon.

A corpus should be large in order to model statistically a maximum of reliable constructions. The more a corpus is important, the better the events are modeled [5]. For machine translation or speech recognition, it is not surprising to train the language model on a corpus of more than 300 million of words.

Classical n-grams models are often enriched by language models based on triggers which are used in several domains, for example in translation, they are exploited to build multilingual dictionaries [19].

The triggers focus on terms that often appear together. That is to say, a word w_i will probably trigger the word w_j . That means we can predict the word w_j when w_i occurs. For instance the word *plane* will probably predict the word *travel*.

The triggers are determined by calculating for each vocabulary word its Mutual Information with each word in the dictionary. Then, only words with a high mutual information are kept used as triggered words [5].

The mutual information is a measure of distance stemming from the information theory, which allows to measure the degree of association between two events. The mutual information $MI(x,y)$ represents the importance of the link which exists between two events x and y . The non-weighted MI is given below:

$$MI(x, y) = \log \frac{P(x, y)}{P(x)P(y)}$$

Where:

$P(x)$ is the marginal probability of x

$P(y)$ is the marginal probability of y

$P(x, y)$ is joint probability of x, y

We propose to enrich each concept (proposed by an expert) of the ontology. For each term we seek in the e-mail corpus the words which are correlated to it, so the terms having an important mutual information are used in order to represent the concept.

The calculation of the appearance of the couple (x,y) is made with regard to a moving window of n terms. The window is determined according to the type of the corpus, for example for a corpus of e-mails, the window has been set to an entire e-mail whereas for a newspaper corpus the window is the paragraph.

3.3 An overview of Training Material

The corpus used is built from e-mails exchanged between the team members in charge of mediatisation of more than 200 educational contents. A team comprises about 60 members, among mediatisors, coordinators and co-authors.

The corpus is recovered from the e-mails of the main coordinators (educational and technical) because they receive all the e-mails exchanged between the team members. Sending e-mails to several persons allows to inform all the team about the evolution of the work, and to reduce the number of meetings.

Table 1. The e-mail corpus

Corpus	71826 words
Vocabulary	4321 words

The corpus of e-mails has the advantage to deal exactly with our purpose but unfortunately it is not huge. That is why, we decided to use a second corpus to enrich our ontology, it is extracted from the Newspaper "Le Monde".

Table 2. The newspaper corpus

Corpus	972.436 words
Vocabulary	43.297 words

Table 1 and 2 show respectively the size of e-mails and newspaper corpora.

Before retrieving triggers the corpus has been cleaned. In particular we remove data which provides no information and introduce noise. The goal is to search triggers among the terms that may represent topics.

We start the process by separating the e-mail body of the attachments. We only work on the e-mail body, because attachments are pedagogical reports. They contain no information on the emotions of the team members, so including them to the corpus reduce the importance of expressed emotions.

Then we remove automatically all the microadvertising (it represents all the forms of publicity which we find on Web pages), figure 3 shows an example of microadvertising in an e-mail.

Do You Yahoo !?

Do away with spam? Mail offers the best protection against unsolicited messages

<http://mail.yahoo.com> Yahoo ! Mail

Figure 3. An example of microadvertising

The third operation of the process concerns the lemmatization of the corpus using TreeTagger² [21], this tool associates with each word a morphological-syntactic label and gives its lemma. It has been developed at the Institute for Computational Linguistics of the University of Stuttgart. It was used for several languages namely German, English, French, Italian, Dutch, Spanish, Bulgarian, Portuguese, etc. The lemmatization allows to find the root of words, what entails an important reduction of the lexicon and an increase of the frequencies of the lemmas which compose it.

The last operation of cleaning concerns the remove of tool words, in order to filter and identify the important words that are meaning-carrying. The tool words may include addresses of web pages (URL: Uniform Resource Locator), e-mails addresses, punctuation, conjunctions, articles, symbols (#, *, &, @, %, +, /, -, etc.), numbers and dates (written in figures or letters).

² <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

4. EVALUATION

We organized our experiments on two steps. The first one aims to study the interpretation of the mutual information, the second consists in looking for triggers in the corpora from the terms of the ontology.

First we sought triggers in the e-mail corpus, but we treated the terms which do not express conflict. The e-mail corpus deals conflicts but the topic which dominates is the mediatization of the educational contents. We chose of the lexicon of the mediatization of the educational contents, the terms: "lesson", "to mediatize" and "qualification" (the latter represents the check of the mediatization by the coordinators and the co-authors).

The term "lesson" trigger the other two terms, but it is much more related to "mediatize" (MI (lesson, mediatize) = 3.29) than to the term "qualification" (MI (lesson, qualification) = 3.17), although the occurrence frequency of the term "qualification" in the corpus is more important than the occurrence frequency of the term "mediatize". In other words, even if the term "mediatize" appears less in the corpus, it appears more with the word "lesson".

Then we sought conflict triggers in both corpora, the corpus of e-mails gave several conflict triggers as: to laugh, problem, complaint, serious, stupid, invent, idiot, horror, etc. However the results are not rich because the corpus is not important. This corpus allowed to find terms which express conflicts indirectly as the term "sir", it is a neutral term and it does not belong to any topic, but according to the cultural point of view it is linked to the conflict topic because it is used in all e-mails of conflict, and several times in the same e-mail.

In the newspaper corpus we found several triggers, for example:

- The term "disappointing" may trigger the term "treason" (MI(disappointing, treason) = 7.09) and the term "crying" (MI(disappointing, crying) = 7.05), etc.
- The term "dissatisfied" may trigger the term "anguish" (MI(dissatisfied, anguish) = 7.88), and the term "to swallow" (MI(dissatisfied, swallow) = 8.52), etc.
- The term "sad" may trigger the term "unhappy" (MI(sad, unhappy) = 7.69), and the term "misunderstanding" (MI(sad, misunderstanding) = 7.38), etc.

These triggers will be incorporated into the ontology to better represent its concepts, in order to detect conflict e-mails. They will also make it possible to assign degrees of importance to conflicts e-mails, such as an e-mail that includes the term "misunderstanding" may have a low degree of importance, because it may indicate that the conflict is not serious and that the concerned manage it well.

Our system has allowed to test and validate the importance of triggers in the search for the semantic links between terms, however, the results depend heavily on the type of corpus used.

5. CONCLUSION

This paper deals with building an ontology of conflicts, in two steps. The first one sets the first-draft of the ontology. Concepts used have been inspired specially from the classification of Michelle Larivey. The second step enriches the concepts of ontology by extracting all the terms that are related to the domain of conflict. For that we used two corpora, the first is a corpus of e-mails, it deals with the subject of conflict but it is relatively small,

the second is a newspaper corpus, its size is quite large, however, it evens out several themes, accordingly results obtained are not rich. So it is necessary to test our system on a substantial corpus and especially where the theme of conflict dominates. To achieve that we will increase our corpus by exploiting the richness of Internet. Each term of the ontology will be triggered by the correlated words coming from the Web.

6. REFERENCES

- [1] Broches, R.S. *Unraveling the Hawthorne Effect: An Experimental Artifact 'Too Good to Die'*. Phd thesis, University of Wesleyan, Connecticut, 2008
- [2] Dias, G. *Extraction automatique d'associations lexicales à partir de corpora*. Phd thesis of University Orléans, 2002.
- [3] McGuinness, D.L. Ontologies come of age, in Fensel, D., Hendler, J., Lieberman, H., Wahlster, W. (Eds), *Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential*, MIT Press, Cambridge, MA, pp.171-92, 2002.
- [4] Kyng, M. Designing for cooperation: cooperating in design, *Communications of the ACM*, vol. 34, no. 12, 65- 73, 1991.
- [5] Haton, J.P., Cerisara, C., Fohr, D., Laprie, Y. Smaili, K. Reconnaissance automatique de la parole. Du signal à son interprétation. *Edition Dunod*. ISBN 210005842 8, 2006.
- [6] Gruber, T. R. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993
- [7] Denoual, E. *Méthodes en caractères pour le traitement automatique des langues*. Phd thesis, University of Joseph Fourier, Grenoble, 2006.
- [8] Haddad, M.H. *Extraction et Impact des connaissances sur les performances des Systèmes de Recherche d'Information*. Phd thesis, University of Joseph Fourier, Grenoble, 2002.
- [9] Aussenac-Gilles, N., Biébow, B., Szulman, S., Terminae Workshop on Evaluation of Ontology Engineering Environments, *13th International Conference, EKAW'2002*, Siguenza, France, CEUR-WS Vol-62, 2002
- [10] Voltz R., Oberle D., Staab S., Motik B. Kaon server - a semantic web management system. *In Alternate Track Proceedings of the Twelfth International World Wide Web Conference*, p. 139-148, Budapest, Hungary : ACM. 20-24 May 2003.
- [11] Smadja, F., Retrieving Collocations from Text : Xtract, In *Computational Linguistics*, no 19(1), pp 143-178, 1993
- [12] Velardi P., Missikof M., Fabriani P. Using text processing techniques to automatically enrich a domain ontology. *In Proceeding of ACMFOIS*, 2001.
- [13] Debois, N. Les émotions en EPS : comprendre et intervenir, éd. *Revue EP.S*, 1 vol. p 77, ISBN 9782867133510, 2007
- [14] Larivey, M. Les genres d'émotions, *La lettre du psy*, vol, no 7/7/1998
- [15] Bourigault, D. LEXTER, *Un logiciel d'extraction de terminologie. Application à l'acquisition des connaissances à partir de textes*. Phd thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris, 1994
- [16] Bourigault, D., Fabre C. Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de grammaire*, 25, 131-151, 2002.
- [17] Daille, B. Study and implementation of combined techniques for automatic extraction of terminology. The Balancing Act : Combining Symbolic and Statistical Approaches to Language, *Proceedings of the Workshop of the 32nd Annual Meeting of the ACL*, 1994, Las Cruces, New Mexico, USA, 9p.
- [18] Enguehard C., Pantera L. Automatic natural acquisition of a terminology. *Journal of Quantitative Linguistics*, 2(1), 27-32, 1995.
- [19] Lavecchia, C., Smaïli K., Langlois D., Haton J.-P. Using inter-lingual triggers for Machine translation, *Eighth conference INTERSPEECH*, 2007
- [20] Jelinek, F., Mercer, L.R. "Interpolated Estimation of Markov Source Parameters from Sparse Data", *Pattern recognition in practice*, in E. S. Gelsema and L. N. Kanal, PP 381-397, 1980
- [21] Kanso, H., Elhore, A., Soule-dupuy, C., Tazi, S. Semi-Automatic Analyzer to Detect Authorial Intentions in Scientific Documents, *International Journal of Computer Science*, vol. 3, no. 1