Automatic Detection of Xenophobic Narratives: A Case Study on Swedish Alternative Media

Lisa Kaati and Amendra Shrestha
Department of Information Technology
Uppsala University
Uppsala, Sweden
firstname.lastname@it.uu.se

Katie Cohen
Department of Psychology
Stockholm University
Stockholm, Sweden
ka_cohen@yahoo.se

Sinna Lindquist
FOI
Swedish Defence Research Agency
Stockholm, Sweden
sinna.lindquist@foi.se

Abstract—In this work we use text analysis to analyze communication on a set of Swedish immigration critic alternative media sites. Our analysis is focused on detecting narratives containing xenophobic and conspiratorial stereotypes. We are also interested in identifying differences in emotional tone and pronoun use in a comparison with traditional media. For our analysis we have used the text analysis tool LIWC (Linguistic Inquiry and Word Count) and a set of dictionaries made to capture a xenophobic narrative. The results show that there are significant differences between regular media and immigration critic alternative media when it comes to the use of narratives and also in the emotional tone and pronoun use.

I. INTRODUCTION

Alternative media are media that differ from established or dominant types of media in terms of their content, how they are produced, or how they are distributed. Alternative media is not a new phenomenon (e.g. the underground press in the 1960s or punk fanzines in the 1970s), and it is more than just printed material, exemplified by broadcasted sound and picture from pirate radio and TV stations [1]. Alternative media and alternative news can be understood as an important part of a democratic society where the free speech and the free press are corner stones. Since not all people have had access to the technology to print, broadcast or publish, alternative media has historically not been democratic in its broadest sense. However, with the Internet and the ready at hand applications to assemble texts, graphics and sound that almost everyone with a computer can use, alternative media nowadays has an important position as a democratic medium [2].

In Sweden, as well as in several other European countries, there has been a recent surge in activity and formation of extreme right movements. These groups show a high inter-activity on forums and blogs, using the Internet and social media as a means for recruiting and spreading their views. In this work we will focus on how these views are presented in immigration critic alternative media. Whether they are nationalist, islamophobic or national-socialist, their common denominators are xenophobia and conspiracism. Hereafter we will use the term alternative media when we refer to the group of alternative media that is critical towards immigration.

It is difficult to know what impact alternative media has on society. However, alternative media is part of the public discourse, especially regarding immigration and critique of traditional media, which show the need for getting a better understanding of this phenomenon. In an interview study with active representatives from a set of Swedish immigration critic alternative media, the aim was to better understand how they perceive traditional journalistic media and how they reason on their own activity in relation to the traditional media, the motivation and the relation between the different alternative media sites, in order to understand the critique of the established media [3]. The motivating factors differs between the interviewees, but has to do with the idea of political correctness that the interviewees find that they have to relate to, and that is maintained by traditional mainstream-media. They all claim that traditional media deliberately use authorized facts and figures to cover-up "the truth" and that the interviewees want to show another perspective, another truth. The critique of established media is noticeable among all interviewees, especially on aspects on immigration and how that is depicted in traditional media. The interviewees claim that the name alternative media works well in terms of getting their ideas out, but are at the same time critical to the term as the sites often are the clattered together with more extremist alternative media sites and labelled "hate media" or "aggression media".

Generally immigration critic alternative media, present themselves as free and independent news sites, where the truth is expressed unfiltered, without regard for political correctness. Simultaneously, they present the established mainstream news sites and information channels as unreliable due to self-censoring and political correctness, and presumably governed by political forces. In [4] a quantitative content analysis of Swedish immigration critic alternative media sites was conducted in order to describe the content in respective media based on themes that has been used to compare different media and news' political relevance. The themes are: news topics, dimension (how often different opinions were represented around one topic), focus (how the news are depicted and put in a larger context with focus on individual or society aspects), style (how the texts are written). The study consisted of 787 articles that were published during 2014 and each article were coded according to the themes. The results show that alternative media is not an alternative in that it could compete or replace the established media when it comes to

121

TABLE I
THE MIXED STEREOTYPE CONTENT MODEL

		Competence	
		Low	High
Warmth	High	**Paternalistic stereotype** low status, not competitive (eg., housewives, elderly people, disabled people)	**Admiration** high status, not competitive (eg., ingroup, close allies)
	Low	**Contemptuous stereotype** low status, competitive (eg., welfare recipients, poor people)	**Envious stereotype** high status, competitive (eg., Asians, Jews, rich people, feminists)

news topics. Alternative media has a clear focus on politics and society issues, crime and war and conflicts. Their texts are more personal and critically positioned and often more negative than in established media, and the vast majority of the editors are men (only 7% were written by women). The same study also did a manual coding of the emotionality/sentiment of the articles and concluded that overall the emotionality is negative with a high variance within the set of alternative media that was studied.

The study show that there are major differences between alternative media sites, mainly shown in if the articles were classified negative, neutral or positive and also in different styles of writing. That indicates that alternative media should not be classified as one type of media.

In this work, we have focused on a set of Swedish alternative media websites and the articles they have published during 2015. The alternative media sites considered all have connections to different extreme right groups. Whether they are nationalist, islamophobic or national-socialist, their common denominators are xenophobia and conspiracism.

The aim of this work is twofold. Firstly, using computerized text analysis, we want to explore how six different Swedish alternative media sites reproduce a far-right narrative identified by experts on Swedish far-right movements. Secondly, with regards to this narrative, as well as other linguistic features such as emotionality, we aim to compare alternative media to traditional mainstream media (hereafter denoted as regular media). For comparison, we have used the websites of Sweden's largest quality press newspaper and Sweden's largest popular press newspaper.

A. Outline

This paper is outlined as follows: Section II describes the method that we have used in this work. In Section III the datasets that we have used in our analysis are presented. In Section IV we present the results of our analysis. Section V provides as discussion about the results. Finally, some conclusions and directions for future work are presented in Section VI.

II. NARRATIVES AND STEREOTYPES

According to [5] , social identity is established, among other ways, by the co-narrating of a story. A shared narrative, i.e. a shared perception of political and historical reality, is essential for cohesiveness in a political group. A common extreme right narrative is a story of conspiracy, starring three groups: The Elite, the People and the Minority. The plot is that in order to gain or preserve power, the Elite uses the Minority against the people. For instance the elite (liberal politicians and Jewish-owned media) imposes minorities (immigration, multiculturalism) on the people ("real people", Swedish workers, elderly, children), while lying to them about the true consequences of immigration (Expo Foundation 2016, personal communication, 3 March, 2016). It follows that anyone who accepts this narrative as true, is able to construe oneself as a person who has called the bluff and gained insight into the real state of affairs.

Each of these characters can be fitted into the stereotype content model (SCM) [6], a well-established psychological theory often used to explain how prejudices are based on relations of competition and status. According to the SCM, perceptions of social groups can be divided into four kinds of stereotypes based on two dimensions: warmth and competence. Table I shows the stereotype content model.

Groups who are not considered to be competing over resources with the ingroup are usually regarded with more warmth than those who are, whereas groups with higher status are considered more competent. These two dimensions form a matrix with three different kinds of outgroups and one ingroup. Groups perceived as incompetent and warm (elderly, disabled, *the People*) fit into the paternalistic stereotype. Groups perceived low in both competence and warmth (welfare recipients, people belonging to "primitive" cultures, *the Minority*) fall into the contemptuous stereotype. The combination of high perceived competence and low warmth (feminists, jews, *the Elite*) constitutes the envious stereotype, whereas high competence and warmth characterizes the group one is part of, or wants to be part of: The admired stereotype.

The words used in the extreme right narrative, as identified by domain experts, also fit the stereotype content model. The Minority is referred to in derogatory terms, with expressions of contempt like "welfare-tourists" or "locusts". The Elite seems to be perceived as aggressive "national state-haters", often referred to with words alluding to establishment and censorship. The People are discussed in a more positive tone ("real people", "community").

III. METHOD

Identifying narratives using text analysis is a difficult task. In this work we use a simple dictionary based approach to find references to the stereotypes that the narrative is based upon. In the first step experienced domain experts manually made dictionaries of words used among the extreme right to refer to any of the three groups. In the next step, the relative frequencies of dictionary words in the different media were calculated by a word count program.

We also run the texts through the word-count program Linguistic Inquiry and Word Count (LIWC) [7], in order to examine if there were any significant differences in any of the existing categories. We specifically wanted to see if there were differences between the alternative and regular media with respect to emotional tone and pronoun use, since elevated use of negative emotion-words and third person plural words are features that have been observed in extremist sites, but not in regular media [8]. LIWC is developed by James W. Pennebaker and it has been evaluated and tested in a number of different studies [9], [8]. LIWC is used to sort words into psychologically meaningful categories. The categories that we consider in this analysis are described in Table II. In our analysis, the relative frequency of each word in a category is computed and the results describes the percentage of how much the words from a category is used.

IV. DATASET

The dataset we have collected consist of articles published on online magazines/websites during 2015. The set of alternative media sites that we consider are critical towards immigration and they were identified by domain experts. For the alternative media, all articles published during 2015 were collected. For the two other media sites (Aftonbladet and DN) we have only collected a sample of articles that were published during 2015. The different media sources we have used in our analysis and the number of articles collected from each source are described below.

Nordfront (619 articles)

Nordfront is the Swedish news and information channel for the Nordic Resistance Movement, a national socialist organization. The site is described as "a news site that stands out and dares to challenge the old, often Jewish owned and/or controlled, mass media empires, free from the censorship and lies that characterizes aforementioned media". It is updated several times a day with selected news characterized by extremism and an underdog perspective, racism and reactionary ideas.

Avpixlat (4391 articles)

Avpixlat, meaning "no pixels", not retouched, is a name referring to the anonymization that politically correct media do to photos to protect criminals, preferably criminals with non-white appearance. According to the Avpixlat site, it expresses a reaction against "old media" that in a conspiratorial way, together with those in power and the elite

TABLE II
THE DIFFERENT CATEGORIES WE FOCUS ON OUR ANALYSIS AND SOME EXAMPLE WORDS.

Categories	Example
Language variables	
3rd person plural	they, their, them
Negative Emotions	hate, worthless, enemy, hurt
Narratives	
Elite	race mixers, anti-swedes
Minority	luxury immigrants, occupants
People	nation, people of reality, Swedes

in the society, selects the reporting on and distorts facts about immigration and crime.

Exponerat (6239 articles)

Exponerat - om sanningen ska fram (Exponerat - if truth be told) describes itself as a politically and religious independent news site publishing and commenting on information about immigration politics, crime, corrupt media, the Swedish police and court proceedings and social welfare.

Fria tider (4856 articles)

Fria tider is a right wing populist news site with opinion similar to the Sweden democratic party and especially to the YoungSwedes (Ungsvenskarna SDU, Sweden Democratic Youth party). Fria tider is critical to the Swedish immigration policy and to the "politically correct old media" and claims to be independent.

Nyheter idag (1400 articles)

Nyheter idag is, as stated on the website, an independent news site focusing reporting on politics, gossip, social media and international news. It is not officially related to any political party. The website is presented as based on a liberal (libertarian) idea of freedom for individuals as well as freedom for information.

Samtiden (2632 articles)

Samtiden present itself as an independent news site with social conservative ideas on editorial page. The site want to present objective news with relevant analysis and a conservative discussion.

DN (1747 articles)

Dagens nyheter (DN) is the largest morning newspaper and quality press in Sweden. It is distributed daily on paper and available online.

Aftonbladet (613 articles)

Aftonbladet is Sweden's largest daily evening newspaper popular press newspaper distributed on the internet and on paper.

V. RESULTS

In this section we will provide the results of our analyses along with some explanations. We have compared the combined results from regular media (Aftonbladet and DN) with
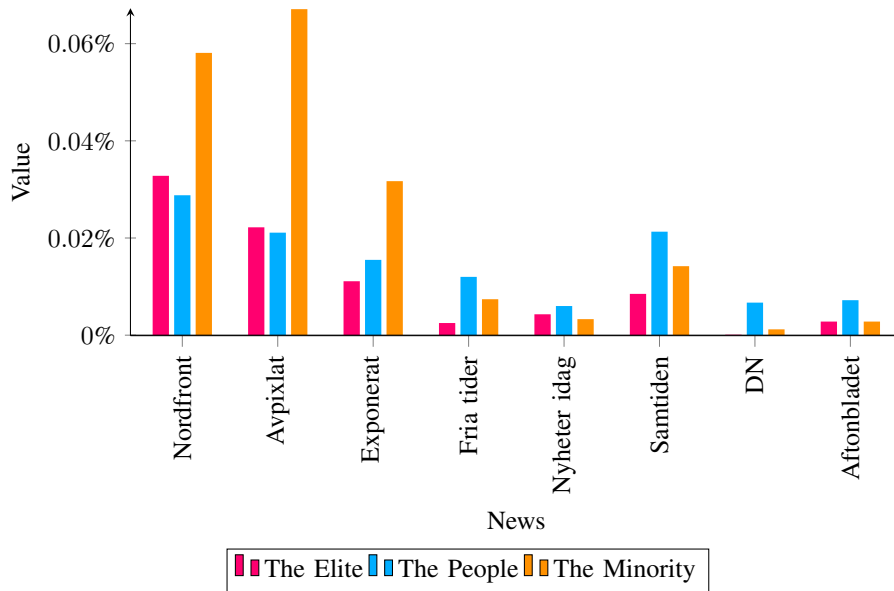
Fig. 1. The use of narratives.

the combined results from alternative media (Avpixlat, Exponerat, Fria tider, Nordfront, Nyheter idag and Samtiden). Table IV shows the results for the different categories (expressed as percent of the total word count), as well as the means for each category for regular and alternative media respectively. Figure 1 shows the frequencies of different stereotype or extremist narrative words in the different media sources.

To verify that there are statistically significant differences between the alternative media sites and the regular media, we have divided the data into two groups: one containing all the alternative media sites and one containing DN and Aftonbladet. The results from a t-test on the two different groups can be seen in Table III. Table V and Table VI shows the results of a t-test on an individual level comparing each alternative media sites with DN and Aftonbladet. In the tables, $*$ indicates a significant difference with $p < 0.05$ while $-$ indicates that there is no significant difference.

TABLE III
SIGNIFICANCE TEST FOR ALTERNATIVE MEDIA COMPARED WITH
NON-ALTERNATIVE MEDIAS.

Category	p-value	Significant
The Elite	7.809190e-13	Yes
The People	4.388677e-03	Yes
The Minority	8.649609e-26	Yes
Negative emotion	2.683428e-03	Yes
Third person plural	7.961904e-03	Yes

When comparing the two different kinds of media we found significant differences regarding how the main characters of the narrative are referred to. Figure 1 shows how words from the three different categories are used in the different media. As can be noted, DN and Aftonbladet uses words from the People category, but hardly anything from the other categories.
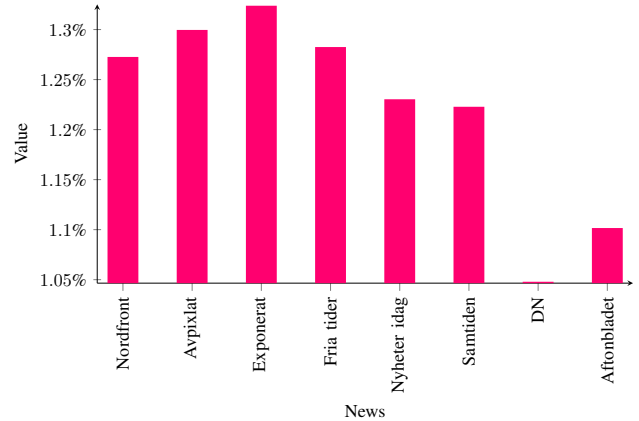

Fig. 2. The use of negative emotions.

Nordfront, Avpixlat and Exponerat uses these words much more frequently than do any of the other sites.

A. Emotionality

Positive emotion words (happy, pretty, good) are used to describe positive events, while negative emotion words (hate, worthless, enemy) are used to describe negative events. Usually, when a person uses a high degree of emotion words in the description of an event, the person is also highly immersed in that event. The emotional tone in extremist material has been studied by Pennebaker and Chung in [8], who found that communication from al Qaida leaders was relatively high in emotion compared to other texts, and also had a higher ratio of negative emotion words compared to positive emotion ones than what is usually found in natural conversation. While natural conversation contains almost twice as many positive

TABLE IV
THE RESULTS FOR THE CATEGORIES WE HAVE USED WHEN COMPARING THE TWO GROUPS OF MEDIA.

	Aftonbladet	DN	Avpixlat	Exponerat	Fria tider	Nordfront	Nyheter idag	Samtiden	Mean reg. media	Mean alt. media
Elite	0.0027	0.0001	0.0222	0.0111	0.0025	0.0329	0.0043	0.0085	0.0014	0.0136
People	0.0069	0.0067	0.0211	0.0155	0.012	0.0288	0.006	0.0213	0.0068	0.0175
Minority	0.0027	0.0012	0.0671	0.0317	0.0074	0.0582	0.0033	0.0142	0.002	0.0303
Negative emotion	1.1016	1.0479	1.2996	1.3238	1.2825	1.2726	1.2303	1.2228	1.075	1.272
Positive emotion	1.7477	1.5934	1.4847	1.3986	1.3223	1.786	1.4571	1.5619	1.671	1.502
Third person plural	1.137	1.0712	1.2537	1.3488	1.1996	1.1278	1.1265	1.2435	1.1041	1.21665

TABLE V
SIGNIFICANCE TEST FOR AFTONBLADET COMPARED WITH ALTERNATIVE MEDIAS.

	Avpixlat	Exponerat	Fria tider	Nordfront	Nyheter idag	Samtiden
Eliten	*	*	-	*	-	-
The People	-	-	-	-	-	-
The Minority	*	*	*	*	-	*
Negative emotion	-	-	*	-	-	-
Third person plural	-	-	-	-	-	-

TABLE VI
SIGNIFICANCE TEST FOR DN COMPARED WITH ALTERNATIVE MEDIAS.

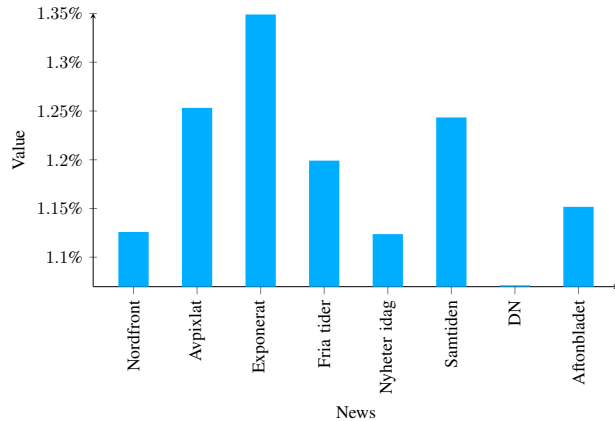|	Avpixlat	Exponerat	Fria tider	Nordfront	Nyheter idag	Samtiden
Eliten	*	*	*	*	-	*
The People	*	*	-	-	-	*
The Minority	*	*	-	*	-	*
Negative emotion	*	*	*	-	-	-
Third person plural	*	*	*	-	*	*

Fig. 3. The use of third person plural.

than negative emotion words, the al Qaida-texts had a much higher relative degree of negative emotion words. We found that the alternative and the regular media were equally emotional overall (Regular media = 2.74, alternative media = 2.77), but that alternative media had a significantly higher frequency of negative emotion words (1.27, 1.07). Correspondingly, the alternative media had a lower frequency of positive emotion words than regular media (1.67, 1,50). Figure 2 shows the frequency of negative emotion words in the different media sources.

B. The use of third person plural

Pronouns are probably the most researched word class in the interdisciplinary field between psychology and linguistics. The use of personal pronouns can be linked to personality, emotions and attitudes[10]. We have chosen to examine the use of third person plural (they, them etc), since it is stated in [8] that a frequent use of third person plural within a group indicates a preoccupation with an oppositional group, or an outgroup. The existence of such an oppositional group serves to strengthen group identity and adherence to the group's cause. According to [8], the use of third person plural in online groups such as American Nazis and animal rights groups has been proven to be the best single predictor of extremism. Figure 3 shows the use of third person plural in all examined media sources. Compared to DN, all alternative media used much more "they"-words, but compared to Aftonbladet only four out of six used more "they"-words. On group level, the alternative media had a significantly higher frequency of third person plural than did the regular media.

VI. DISCUSSION

Overall, immigration critic alternative media is more negative than in regular media. This result is also supported by [4] in a study were two humans manually coded 787 articles from alternative media (from year 2014).

There is a higher use of third person plural on the alternative media sites, which can be a sign of outgroup-focus, or even xenophobia. However, it can also be regarded as an artifact of

the focus on immigration and immigrants that prevails in all the alternative media considered here. The alternative media does not, as regular media, set out to cover as much of current events as possible, but rather selects certain topics of interest. Thus, the coverage in alternative media constitutes a subset of the coverage in regular media, suggesting that the comparison might not be completely fair. This is also described in [4] were it is concluded that alternative media has a clear focus on the topics politics and society issues, crime and war and conflicts.

Looking at each alternative media site individually, we can observe a variance in the frequency of words referring to the characters in the extremist right narrative. This indicates that that alternative media should not be classified as one type of media, something that was also noticed [4].

VII. CONCLUSION AND FUTURE WORK

In this work we have used text analysis to detect narratives in articles published in a set of alternative media news sites, and also to detect differences between articles published at the set of alternative media sites and articles published in regular media (represented by Sweden's largest quality and popular paper respectively). We found that the narratives identified by experts on extreme right organizations were built on group descriptions that fitted well into the stereotype content model. We then tested to what extent these stereotypes prevailed on the different media sources, and found that they were present in alternative media and almost non-existent in the regular media. We also used LIWC to compare the articles published by regular media to the ones published by alternative media, and found that negative emotion words and third person plural were significantly more frequent in alternative media than in regular media.

We conclude that since stereotype words are vastly more common in the examined alternative media than in regular media, the extreme right narrative is reproduced to a much higher extent in the alternative media than in regular media. The fact that third person plural was used significantly more in the alternative media sources supports the notion that the examined alternative media emphasizes the contrast between ingroup and outgroup [8]. We also conclude that the relatively high frequency of negative emotion words and low frequency of positive emotion words in alternative media indicates that a higher degree of discontent is communicated on the alternative media. This is consistent with the manual content analysis mentioned in [4] that found the alternative media more negative than regular media.

Our findings show the possibility of using computerized methods of text analysis to find bias, stereotypes or extremism in online material. For future work, more advanced methods, for instance machine learning, should be used for detecting narratives. Dictionary-based approaches, like LIWC and manually composed dictionaries, do have the drawback of being static, and not able to capture certain linguistic features such as irony or metaphors. Using semantic technologies such as word2vec [11] and random indexing [12] to detect synonyms

to the words in each category, the dictionaries can be made more dynamic, and less dependent on human analysis.

REFERENCES

[1] C. Atton, *Alternative Media.* Thousand Oaks: Sage Publications, Inc., 2002.

[2] J. F. Downing, *Radical Media. Rebellious Communication and Social movements.* Thousand Oaks: Sage Publications, Inc., 2001.

[3] K. Holt, ""alternativemedier"? en intervjustudie om mediekritik och mediemisstro," in *In Migrationen i medierna – men det får en väl inte prata om? Institutet för mediestudier*, 2016.

[4] K. Holt, "Skilda verkligheter? "internets undervegetation" vs "pk-maffian"," in *In Migrationen i medierna – men det får en väl inte prata om? Institutet för mediestudier*, 2016.

[5] E. Ochs, "Constructing social identity: A language socialization perspective," *Research on language and social interaction*, vol. 26, no. 3, pp. 287–306, 1993.

[6] S. T. Fiske, C. A. J., P. Glick, and J. Xu, "A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition," *Journal of personality and social psychology*, vol. 82, no. 6, 2002.

[7] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count (liwc): A text analysis program." New York: Erlbaum Publishers, 2001.

[8] J. W. Pennebaker and C. K. Chung, "Computerized text analysis of al-Qaeda transcripts," in *The Content Analysis Reader*, K. Krippendorf and M. A. Bock, Eds. Sage, 2008.

[9] Y. Tausczik and P. J.W., "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, March 2010.

[10] J. W. Pennebaker, *The secret life of pronouns: What our words say about us.* CT New York: Bloomsbury Press., 2011.

[11] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," *ArXiv e-prints*, Feb. 2014.

[12] M. Sahlgren, "An introduction to random indexing," in *Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering, TKE 2005*, 2005.