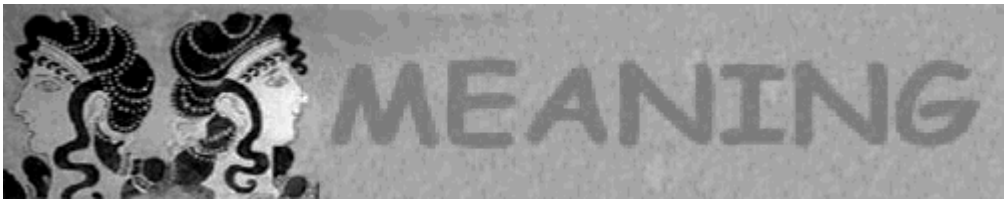
MEANING-full effects on information retrieval

03/03/2005

Version 2



Developing Multilingual Web-scale language Technologies

INFORMATION SOCIETY TECHNOLOGIES

Deliverable 8.3.

Work package 8, Validation and demonstration

Project ref.	IST-2001-34460
Project Acronim	MEANING
Project full title	Developing Multilingual web-scale language technologies
Security (distribution level)	Public
Contractual date of delivery	
Actual date of delivery	
Document number	D8.3
Type	Report
WP contributing to the deliverable	WP8
WP task responsible	German Rigau
EC project officer	Evangelia Markidou
Authors	Piek Vossen, Elton Glaser, Mihaela Gradinaru, Rachel van Steenwijk, Hetty van Zutphen,
Keywords	Human Language Technology, Cross-lingual Information Retrieval, Natural Language Processing, WSD, large-scale knowledge bases.
Abstract	This deliverable reports on testing the use and effect of the integration of the MEANING technology in the TwentyOne search engine of Irion.

Table of Contents

1 Introduction 4

2 TwentyOne Search Indexes 5

3 Queries 8

4 Results 13

5 Conclusions 17

6 References 19

1 Introduction

This deliverable describes the validation of MEANING in a cross-lingual Information Retrieval application. MEANING acquires lexical knowledge from various sources and various languages. This knowledge is stored in the Multilingual Central Repository (MCR), which is based on the design of the EuroWordNet database. The MCR holds wordnets in various languages (English, Spanish, Italian, Catalan and Basque), which are interconnected via an Inter-Lingual-Index (ILI). MEANING uses WordNet1.6 as an ILI to share lexical knowledge stored for each separate wordnet. During the MEANING project, the MCR has been enriched in various cycles. The purpose of work package 8 is to demonstrate that the results of MEANING can be integrated in a real application and on real data.

In MEANING Deliverable 8.2, we explained how the MEANING results have been integrated in the Irion TwentyOne Search Engine and have been applied to a collection of Spanish and English captions with pictures from the Spanish publisher EFE. The Spanish and English indexes have been expanded with the Basque, Italian and Catalan wordnets from the MCR. Word-Sense-Disambiguation was done using Wordnet Domains from the MCR. More details are explained in Deliverable 8.2. We assume that you have read Deliverable 8.2 before reading this deliverable.

In this deliverable, we describe a cross-lingual retrieval experiment on these indexes. For this experiment, we used the TwentyOne retrieval benchmarking environment. This environment lets you run a collection of queries on a TwentyOne index and automatically derives the recall figures. The queries can be formulated in any of the search languages, in this case: Spanish, English, Basque, Italian and Catalan.

The deliverable first describes the TwentyOne Search system and the 3 indexes that have been built (section 2). Section 3 specifies the queries and the test environment and section 4, the results of running the queries on the different indexes.

2 TwentyOne Search Indexes

TwentyOne Search is a conceptual search engine that uses a combination of statistical and language-technology techniques. It is a two step system, where first, the relevant documents are collected using state-of-the-art statistical engines, and secondly, the best matching phrases from the relevant documents are collected. The statistical core-engine of TwentyOne Search returns the most relevant documents from large collections, using a standard vector-space weighting. It ensures fast and robust retrieval. The language-technology has two major roles:

1. Maximize the recall of the statistical engine so that any document is found regardless of the wording and regardless of the query word choice;
2. Maximize the precision by conceptually matching queries with phrases from documents rather than complete documents;

The phrases in documents are labelled as NPs. The TwentyOne system then uses a range of factors to compare the phrase with the query:

1. Number of matching concepts between the query and phrase;
2. Degree of fuzziness mismatch between the query word and the phrase;
3. Degree of derivational mismatch, compounding, etc.;
4. Whether or not a synonym is used;
5. Whether or not the same language is used;

The effect is that documents with phrases (NPs) that include most concepts and have the most similar wording with the query are shown first. The contextual effect of the phrase match is very powerful, as we will see later.

For the experiment, we indexed a collection of XML records from the EFE publisher with captions for pictures. This collection can be used to find pictures using text queries on the captions. Most of the captions were Spanish, about 10% was English (see Deliverable 8.2 for details).

Table 1: Fototeca data from EFE for April-May 2004

		Spanish	English
XML records	29511	26546	2965
Images	29943		

Three different indexes were built for the EFE collection:

1. EFE_NO3: Without using the MEANING wordnets;
2. EFE_FULL3: Using the MEANING wordnets but without word-sense-disambiguation;
3. EFE_MEANING3: Using the MEANING wordnets and using word-sense-disambiguation;

In the case of EFE_NO3, we built indexes for all 5 languages but the original English and Spanish words were simply copied to the indexes for English, Spanish, Catalan, Basque and Italian. No synonym expansion was applied for English and Spanish and no translation for the other languages.

For indexes 2.) and 3.), the Spanish and English indexes have been expanded to synonyms and translated to English (in case of Spanish), Spanish (in case of English), Basque, Italian and Catalan with wordnets from the MCR. In the case of index EFE_FULL3, all the meanings of the words in the articles have been taken and have been expanded to all synonyms and/or translations. In the case of EFE_MEANING3, we first excluded unlikely meanings using a word-sense-disambiguation system and expanded all the remainder queries.

This means that for all three indexes queries can be made in 5 languages: Spanish, English, Basque, Italian and Catalan, while the system returns both English and Spanish articles. However, queries with synonyms and queries in other languages than Spanish (and to some extend English) will be hardly effective in the 1st index EFE_NO3.

The Word-Sense-Disambiguation (WSD) was done using Wordnet Domains (version 1.1.1, Magnini et al 2002) from the MCR. The basic approach, described in detail in D8.2, consist of classifying the articles with a text classifier that is based on the Spanish and English words associated with domain labels. The text classifier first assigns domain labels to the article as a whole, based on the complete content: so-called *microworld* tags. Next, it also classifies the separate NPs within each article using a window of 10 surrounding NPs (4 to the left and 5 to right). This results in one or more so-called *nanoworld* tags for each NP. All domains scoring above 60% confidence are assigned to have sufficient recall.

The disambiguation then consists of the following process for each word in the NP:

1. Are there word meanings with domain labels that match any of the nanoworld tags? If yes, these meanings are selected.
2. If no, are there there word meanings with domain labels that match the microworld tags? If yes these meanings are selected.
3. If no all meanings are selected.

The concept reduction as a result of the disambiguation is very effective. For the microworlds the reduction is about 48% for Spanish and 57% for English. In the case of the nanoworlds, the reduction is even higher: 52% for Spanish and 65% for English. Most of these reductions (about 44%) however relate to the factotum words (Magnini et al 2002).

To be able to build the cross-lingual search system for the 5 languages and to apply the WSD, we had to import 5 wordnets from the MCR into the TwentyOne system and to build a domain classifier for English and Spanish from Wordnet domains that has been included in the MCR. The details on the integration process have been described in deliverable 8.2.

3 Queries

We used the TwentyOne benchmarking environment to extract queries. This environment extracts NPs from the documents and presents them as queries. These queries are stored in an Xml file with the identifiers for the database, the document and the page from which they originate. Also the NP itself is stored. On the next page you can see two examples of queries that have been extracted in this way. The first query is extracted from database EFE_1. It originates from document 13 and page 85. We extracted NP number 11: “Una colisión en cadena”. From this NP, the Spanish query “colisión en cadena” was derived.

The query file can be used to launch queries on the same database and measure the recall, where we measure the number of times that the same page is returned in the top-segment of the results. The size of the top-segment can be varied. In this experiment, we looked at the top 10 results only. Note that the ranking of the pages is based on the conceptual scoring of the TwentyOne system. This means that those pages end in the top that have phrases (NPs) with most concepts represented and also have the closest wording with respect to the query. The page ranking thus reflects the conceptual phrase ranking, i.e. a synonym will be ranked lower than the original, fuzzy matches lower than precise matches, etc. If pages have the same conceptual score, the page relevance score is used to further differentiate results. The page relevance score is based on a classical vector-space ranking of the query words with respect to each document and does not look at phrases.

The extraction of the NPs can be triggered in various ways. It is for example possible to extracts NPs with a minimal or maximal number of words or NPs that need to include at least one word from a set of words. In this case, we searched for NPs that include Spanish words that are interesting from the perspective of WSD. They either show a clear ambiguity and/or have a common synonym. We verified if the other meaning and/or synonym also occurred in the index, for example *estrella* (*star* in all its meanings) or *figura* (*body*, *form, figure, character, statue*). Finally, we also looked at the relevance of the words to the pictures that go with the article. Most of this work was done manually.

```
<TESTIN>
        <DBS_ID>EFE_1</DBS_ID>
        <DOC_ID>13</DOC_ID>
        <PAG_TITLE></PAG_TITLE>
        <PAG_ID>85</PAG_ID>
        <NPS>
        <NP ID="11">Una colisión en cadena</NP>
        </NPS>
        <SOURCE_LNG>es</SOURCE_LNG>
        <BOOLEAN>AND</BOOLEAN>
        <QUERY_EN>chain collision</QUERY_EN>
        <QUERY_CA>colisió cadena</QUERY_CA>
        <QUERY_IT>collisione a catena</QUERY_IT>
        <QUERY_BA>kolpeak bata bestearen segidan</QUERY_BA>
        <QUERY_ES>colisión en cadena</QUERY_ES>
</TESTIN>
<TESTIN>
        <DBS_ID>EFE_1</DBS_ID>
        <DOC_ID>58</DOC_ID>
        <PAG_TITLE></PAG_TITLE>
        <PAG_ID>105</PAG_ID>
        <NPS>
        <NP ID="17">celebrada en el Palacio de Exposiciones de Ginebra. El Pilatus PC12 es un
avión de turbina que puede aterrizar en pistas</NP>
        </NPS>
        <SOURCE_LNG>es</SOURCE_LNG>
        <BOOLEAN>AND</BOOLEAN>
        <QUERY_EN>propeller plane</QUERY_EN>
        <QUERY_CA>avió turbina</QUERY_CA>
        <QUERY_IT>aereo a elica</QUERY_IT>
        <QUERY_BA>helize hegazkina</QUERY_BA>
        <QUERY_ES>avión de turbina</QUERY_ES>
</TESTIN>
```

In total 58 Spanish queries were extracted, where each query consists of two or more words. The context of the words is sometimes disambiguating, sometimes not. Table 2 shows the complete list of the unique Spanish multiword queries.

Table 2: Original Spanish multiword queries

aficionados muestran pertenencias	figura de chocolate	operación quirúrgico
atuendo típico mexicano	figuras de globos	pancarta contra el terrorismo
avión de turbina	flores violetas	pasajeros facturan equipajes
balón firmado	fábrica química	pastores oran por paz
banda de viento	globos de colores	perro ataviado
buque polar	hijas siamesas	perros guías
cartel contra caza de focas	imagen aérea de ciudad	platillos voladores
catación de café	imagen del papa	presos amotinados
chuletas con verduras asadas	imágenes del cerebro	réplica de moneda
chuletas con verduras	incendio en almacén de papel	sacerdote bendice ramos
clavel sobre ataúd	incendio en fábrica química	sepelio en Medellín
cocaína en piñas	insectos nutritivos	siembra de alubia
colisión en cadena	jarro ornamental	supervivientes de ventisca
comida en el muelle	joven enciende velas	torero con vaqueros
damas arreglan vestido	joven toca batería	trompetista cubano
dinosaurio con plumas	marcha de maestros disidentes	vagabunda dormita
diseño de moneda	mosca sobre hoja	verduras asadas
embarcación con imagen Jesús	mujer de pandilla	verduras gratis
ensayo con elefantes	máscara de cuero	
exhibición de arte contemporáneo	nave cósmica	

The results of the queries are not unique. Most of the words occur frequently in the collection. When we found multiple correct results for a query, we multiplied the queries for each result. For the 58 unique strings, this resulted in a total of 105 queries that were posted, where some queries have up to 6 results.

From the multiword queries, we derived single word queries by selecting a word from the multiword query. This is the ambiguous word or the word with a common synonym. In some cases, multiple words were selected from the query. This resulted in a list of 73 unique single word queries, listed in Table 3

Table 3: Derived Spanish singleword queries

aficionados	cría	imagen	pandilla	siamesas
almacén	cuero	imanes	panel	sirena
araña	damas	imágenes	papa	tabla
balón	diseño	imán	papel	tablero
banda	dominó	instrumentos	pasajeros	terraza
batería	embarcación	joven	pastores	típico
buque	ensayo	jugador	pieza	vagabundo
cadena	estrella	maestro	piñas	vaqueros
cadenas	figura	maestros	planta	velas
café	figuras	mascara	platillos	vestido
camello	flores	muelle	platos	vigilante
cartel	fábrica	naranja	plumas	violetas
caza	globos	nave	química	voladores
cerebro	guías	operación	réplica	
clavel	hoja	pancarta	sellos	

The results from the original query word have been associated with the single words as well. Some of the single words thus also have multiple results. This resulted in 92 single word query postings for Spanish.

The multiword and single word queries have been used to derive synonym queries or paraphrases. If one of the query words had a common synonym and the synonym occurred in the collection, we replaced the word by the synonym to derive a synonym query. Using this method 77 unique synonym queries have been made, where the query consists of multiple words. The original results have been applied to the paraphrased queries, which resulted in 94 query postings. The same was applied to the single word queries. This resulted in 69 unique single word queries and 96 query postings.

Next, the original Spanish multiword queries and the derived Spanish single word queries have been translated by native speakers to English, Basque, Catalan and Italian[1]. The single words got the same translation as the same word in the multiword query. Furthermore, if there are multiple results for the original Spanish query, these results have been copied to the translations, resulting in multiple query postings as well.

Cross-lingual translations exhibit a different pattern of polysemy and synonymy. It can very well be that the translation of a Spanish word in Basque is not ambiguous or has no synonym, or the other way around. Likewise, the translations of the queries have different numbers of unique strings across languages but (almost) the same amount of postings.

Table 4: Unique queries and query postings

	Multiwords		Singlewords	
	Unique strings	Postings	Unique strings	Postings
Spanish original	58	105	77	92
Spanish synonym	77	94	69	96
English	57	105	74	92
Catalan	60	105	69	92
Basque	57	104	65	92
Italian	56	105	74	92

Table 4 then lists the final sets of queries that have been applied to the TwentyOne indexes of the EFE data. Each of these sets has been applied to the 3 indexes of the EFE data. The results are described next.

[1] We would like to thank German Rigau, Eneko Agirre and Manuela Speranza for the translations.

4 Results

The use of wordnets to build the indexes will have an effect on the ranking and on the number of results shown. If Spanish and English words are expanded to all synonyms and translations, the recall will be bigger but also more irrelevant results will be generated. Here the fact that TwentyOne finds phrases and not pages or documents will have a big impact. Even though a wordnet may generate many irrelevant synonyms, the chances that combinations of irrelevant synonyms and/or words are also used by a user in a query are extremely low. We thus can say that the phrase-based retrieval will already have a very strong disambiguating effect.

The results of launching the above queries on the 3 indexes are listed in the next two tables. Table 5 lists the results for the multiword queries and Table 6 for the single word queries. Each table lists the results per index on the rows (NO3, FULL3 and MEANING3) and per language or synonyms in the columns. There are three columns per language:

Q = number of query postings

Page = number correct recall in the first 10 results

% = proportion of recall out of all postings

Each index has a row for the total results and three more rows for the 1st, 2nd and 3rd position (p1, p2 and p3) in the result list. We marked the best scores for the totals and for the 1st position (p1). We did not list the other positions from the top 10 because all the results listed the correct match in the top 3 or outside the top 10.

Table 5: Retrieval results for multi word queries

Multiwords	Spanish original			Spanish synonym			English			Catalan			Basque			Italian		
	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%
NO3	105	99	0.94	94	14	0.15	105	2	0.02	105	31	0.3	104	1	0.01	105	3	0.03
p1		60	0.57		9	0.1		0	0		21	0.2		1	0.01		2	0.02
p2		30	0.29		5	0.05		1	0.01		8	0.08		0	0		1	0.01
p3		9	0.09		0	0		1	0.01		2	0.02		0	0		0	0
FULL3	105	96	0.91	94	71	0.76	105	39	0.37	105	70	0.67	104	50	0.48	105	39	0.37
p1		55	0.52		38	0.4		16	0.15		44	0.42		27	0.26		19	0.18
p2		33	0.31		27	0.29		17	0.16		22	0.21		19	0.18		15	0.14
p3		8	0.08		6	0.06		6	0.06		4	0.04		4	0.04		5	0.05
MEANING3	105	97	0.92	94	61	0.65	105	39	0.37	105	68	0.65	104	46	0.44	105	32	0.3
p1		60	0.57		39	0.41		21	0.2		48	0.46		27	0.26		20	0.19
p2		31	0.3		18	0.19		13	0.12		16	0.15		15	0.14		6	0.06
p3		6	0.06		4	0.04		5	0.05		4	0.04		4	0.04		6	0.06

Table 6: Retrieval results for single word queries

Singlewords	Spanish original			Spanish synonym			English			Catalan			Basque			Italian		
	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%	Q	Page	%
NO3	92	13	0.14	96	4	0.04	92	3	0.03	92	8	0.09	92	5	0.05	92	6	0.07
p1		7	0.08		3	0.03		2	0.02		3	0.03		3	0.03		4	0.04
p2		3	0.03		0	0		1	0.01		1	0.01		1	0.01		1	0.01
p3		0	0		1	0.01		0	0		4	0.04		0	0		1	0.01
FULL3	92	17	0.18	96	8	0.08	92	11	0.12	92	14	0.15	92	16	0.17	92	11	0.12
p1		10	0.11		3	0.03		7	0.08		6	0.07		9	0.1		9	0.1
p2		3	0.03		1	0.01		2	0.02		5	0.05		4	0.04		1	0.01
p3		4	0.04		4	0.04		2	0.02		3	0.03		3	0.03		1	0.01
MEANING3	92	18	0.2	96	8	0.08	92	10	0.11	92	11	0.12	92	16	0.17	92	11	0.12
p1		10	0.11		3	0.03		7	0.08		6	0.07		7	0.08		8	0.09
p2		6	0.07		1	0.01		2	0.02		2	0.02		3	0.03		2	0.02
p3		2	0.02		4	0.04		1	0.01		3	0.03		6	0.07		1	0.01

Let us first consider the multiword queries. The first thing to be noticed is the high recall. The best results are for the original Spanish words on the NO3 index (no use of wordnets): 94%. This is inherent to the conceptual phrase search. The search engine will select NPs that include all the query concepts and give preference to NPs that closely match the query. When we do not use wordnets, as in NO3, the most equal phrases are likely to show up first, especially since the queries have been derived from the NPs and there are not that many NPs with all the query words.

We also see that we hardly loose anything when we use wordnets. The fully expanded index (FULL3) scores 91% and the disambiguated index (MEANING3) scores 92%. This is a major difference with respect to the results reported in the Reuters experiments described in D8.1 (Vossen et al 2004). In Reuters, the retrieval was based on the page score and not on the conceptual phrase score. The conceptual phrase matching thus adds precision. So even if the wordnets add more possible hits and more noise, the fact that the closest wordings are preferred selects the most appropriate results. This is also clear when we look at the p1 positions. Here NO3 and MEANING3 score equally well.

When we look at the queries where a synonymous word was used (the 2nd column group), we see that the index without wordnets (NO3) drops to 15% but the FULL3 index only drops to 76% and the MEANING3 index drops to 65%. This clearly shows the usefulness of wordnets for information retrieval. We also see that the WSD apparently removed certain synonyms that are useful, hence the difference of 10% between FULL3 and MEANING3. This indicates that the WSD settings have been too strict (50% of the concepts have been excluded).

On the other hand, if we look at the p1 scores, we see that MEANING3 scores better than FULL3. This means FULL3 generates more noise that is interfering with the correct results for the 1st position but the correct results apparently still end up in the top 10. This also implies that the total results for FULL3 can be worse than MEANING3 if the index is bigger. In a bigger index there is more competition and the noisy result will push correct results out of the top 10. The pattern that we see for the synonyms also shows up for the cross-lingual

retrieval. FULL3 mostly has best results and MEANING3 is very close but scores better for p1. NO3 has dramatically bad results.[2]

The 1st position results can be seen as a measurement of precision. The disambiguated index thus has a better precision than the fully expanded index. These results are confirmed in the end-user evaluation that is described in the MEANING deliverable 8.4. This evaluation showed that MEANING3 and FULL3 both increase the productivity of the end-users that search for pictures in the database. However, MEANING3 significantly outperforms the FULL3 index. Because the 1st result is more often the correct result, the users can quickly and correctly finish their task and do not have to go through all the other results.

The second table gives the results for the single word queries. In the case of a single word query, the phrase matching is almost completely absent. This means that there is no disambiguating effect of the phrase to filter out noise generated by the wordnets. Nevertheless, NO3 does not have the best results when we search with the original word. The best results (both overall and on p1) are now given for MEANING3, closely followed by FULL3. Apparently there is sufficient natural competition that keeps the correct results from the top-10 list. This is probably due to the fact that we deliberately selected ambiguous words. Overall, the recall dropped for all 3 indexes. This means that all 3 indexes suffer from the competition by other meanings and results. This effect will be bigger for bigger indexes. Note that the results of the system can still be good. We simply do not know whether the returned results are good or bad. We only measure if the original page with the NP is returned in the top-10.

When we look at the synonyms and the translations, we see the pattern of the multiword queries repeated. NO3 has dramatic results (3% up to 7%) and both FULL3 and MEANING3 have equally good results (8% up to 17%), sometimes as good as the original words. Again the use of wordnets is evident in these cases but we do no see much difference between FULL3 and MEANING3. Apparently, the noise generated by FULL3 does not lead to lower recall figures for single words. This is a bit surprising since only 50% of the concepts have been selected. On the other hand, the recall figures are already considerably lower than for

[2] Catalan scores almost as good as the Spanish synonyms. This shows that the languages are closely related. The fact that both the wordnets are developed by the same group can also be a factor.

the multiword searches. It may be possible that there is too much competition for these words and the correctly returned cases are unambiguous for both FULL3 and MEANING3.

5 Conclusions

In this deliverable, we described the full integration of MEANING in the TwentyOne Search engine, and how it has been applied to retrieval of the Spanish and English picture captions of the EFE database. We used the MEANING results of the last MEANING cycle, which includes wordnets for Spanish, English, Basque, Catalan and Italian. We also used the latest version of WordNet Domains from the Multilingual Central Repository (MCR).

We built 3 different indexes on the EFE data: one index without using the wordnets, one index using the wordnets and no word-sense-disambiguation and a 3rd index using wordnets and using word-sense-disambiguation. In the index with word-sense-disambiguation 50% of the meanings have been selected. We extracted Spanish queries from the documents, where we looked at words with multiple meanings and common synonyms. These queries have been paraphrased and translated to the other languages.

The experiments showed that the use of wordnets is definitely useful for both monolingual and cross-lingual retrieval. This is apparent when the queries are paraphrases of original multiword expressions in the documents. When the original words are used in the multiword queries, the indexes built with wordnets did not perform significantly worse. When we look at singleword queries, the wordnet indexes even performed better for both the original words and paraphrases.

The experiments did not show a significant improvement for word-sense-disambiguation in overall recall. Overall results for multiword queries are better without word-sense-disambiguation (considering the top 10 results) but when we look at the first position of the result list the index with word-sense-disambiguation performed better. This may point to a better precision due to the selection of meaning by the word-sense-disambiguation. Less noisy results seem to disturb the top-ranking of the results but not the overall ranking. For singleword queries the indexes with and without word-sense-disambiguation perform the same. The precision effect of the disambiguated index has also been confirmed by the end-

user evaluation described in deliverable 8.4. End-users were most productive with the disambiguated index.

These results are significantly better than have been reported for the Reuters news collection in D8.1. This is partly due to the improved retrieval system that uses conceptual phrase search rather than a page search based on vector space weighting. The experiments clearly showed the effectiveness of a proper conceptual phrase matching in document retrieval. The phrase retrieval has a very strong disambiguating effect. We expect that the use of wordnets and of word-sense-disambiguation will be more effective when bigger indexes are built. In a bigger index there will be more competition for recall and likewise word-sense-disambiguation is needed more to reduce the noise. Another reason may be that we now used the MCR data directly, whereas the Reuters indexes were built with our corporate semantic network (SemNet) that was automatically linked to the MCR.

It should be noted that we only implemented a very basic and simple word-sense-disambiguation system. There is plenty of room for a more effective and high-quality word-sense-disambiguation. Not only is the text classifier very unbalanced (some domains are over-trained) but also is it possible to fine-tune the classification process to cut out less meanings and thus have more recall in the disambiguated indexes.

Finally, the experimental setup of the test and the selection of the proper queries are essential in demonstrating the use of wordnets in retrieval. Retrieval of phrases instead of documents is only possible when there is sufficient recall. Furthermore, the use of small queries where we selected specific ambiguous words or words with common synonyms is also crucial for detecting the differences across the different indexes.

6 References

Fellbaum, C. (ed), WordNet. An Electronic Lexical Database, The MIT Press 1998.

Magnini, B. and G Cavagliá, Integrating subject field codes into wordnet. In Proceedings of the Second International Conference on Language Resources and Evaluation LREC'2000, Athens, Greece 2000.

Rigau, G., B. Magnini, E. Agirre, P. Vossen and J. Carroll. Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of COLING Workshop, Taipei, Taiwan, 2002.

Rigau, G and E. Agirre and J. Atserias, The MEANING project, Proceedings of the XIX Congreso de la Sociedad Espagna para el Procesamiento del Lenguaje Natural (SEPLN'03), Alcalte Henares, Spain, 2003.

Vossen, P. (ed) EuroWordNet: A Multilingual Database with Lexical Semantic Networks, Kluwer Academic Publishers, Dordrecht, 1998.

Vossen, P., E. Glaser, H. Van Zutphen, R. Steenwijk, "Validation of MEANING", *WP8.1 Deliverable 8.1,* MEANING, IST-2001-34460, Irion Technologies BV, Delft, The Netherlands. 2004.

Vossen Piek, German Rigau, Iñaki Alegria, Eneko Agirre, David Farwell, Manuel Fuentes. Validation of MEANING2, MEANING, IST-2001-34460, Deliverable 8.2. 2005.