

# University of Hagen at CLEF 2004: Indexing and Translating Concepts for the GIRT Task

Johannes Leveling, Sven Hartrumpf  
Intelligent Information and Communication Systems  
University of Hagen (FernUniversität in Hagen)  
58084 Hagen, Germany  
{*Johannes.Leveling,Sven.Hartrumpf*}@fernuni-hagen.de

## Abstract

This paper describes the work done at the University of Hagen for our participation at the German Indexing and Retrieval Test (GIRT) task of the CLEF 2004 evaluation campaign. We conducted both monolingual and bilingual information retrieval experiments. For monolingual experiments with the German document collection, the focus is on applying and comparing three indexing methods targeting full word forms, disambiguated concepts, and extended semantic networks. The bilingual experiments for retrieving English documents for German topics rely on translating and expanding query terms based on a ranking of semantically related English terms for a German concept. English translations are compiled from heterogeneous resources, including multilingual lexicons such as EuroWordNet and dictionaries available online.

## 1 Introduction

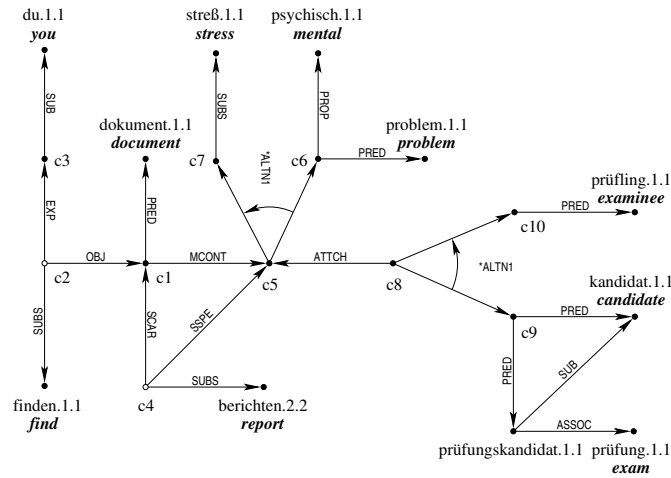
This paper investigates the performance of different indexing methods and automated retrieval strategies for the second participation in the University of Hagen in the evaluation campaign of the Cross Language Evaluation Forum (CLEF). In 2003, retrieval strategies based on generating query variants for a natural language (NL) query were compared (Leveling, 2003). The result of our best experiment in 2003 with respect to mean average precision (MAP) was 0.2064 for a run using both topic title and description.

Before presenting the different approaches and results for monolingual and bilingual retrieval experiments with the GIRT documents (German Indexing and Retrieval Test database, see Kluck and Gey (2001)) in separate sections, a short overview of the analysis and transformation of natural language queries and documents is given. Natural language processing (NLP) as described in the following subsections is part of query processing for the NLI-Z39.50<sup>1</sup> (Leveling and Helbig, 2002), a natural language interface for databases supporting the Internet protocol Z39.50. The major part of our experimental infrastructure was developed for and is applied in this NLP system.

### 1.1 The MultiNet Paradigm

In the NLI-Z39.50, natural language queries (corresponding to a topic's title, description, or narrative) are transformed into a well-documented knowledge and meaning representation, Multilayered Extended Semantic Networks (abbreviated as MultiNet) (Helbig, 2001; Helbig and Gnörlich, 2002). The core of a MultiNet consists of concepts (nodes) and semantic relations and functions between them (edges). Figure 1 shows the relational structure of the MultiNet representation for the description of GIRT topic 116. The MultiNet Paradigm defines a fixed set of 93 semantic relations (plus a set of functions) to describe the meaning connections between concepts, including synonymy (SYNO), subordination, i.e. hyponymy and hypernymy (SUB), meronymy and holonymy (PARS), antonymy (ANTO), and relations for change of

<sup>1</sup>The NLI-Z39.50 is being developed as part of the project "Natürlichsprachliches Interface für die internationale Standardschnittstelle Z39.50" and funded by the DFG (Deutsche Forschungsgemeinschaft) within the support program for libraries "Modernisierung und Rationalisierung in wissenschaftlichen Bibliotheken".



**Figure 1:** The core MultiNet representation for the query “Finde Dokumente, die über psychische Probleme oder Stress von Prüfungskandidaten oder Prüflingen berichten.” / ‘Find documents reporting on mental problems or stress of examination candidates or examinees.’ (description of GIRT topic 116). Numerical suffixes for reading distinction of English concepts are omitted.

sorts between lexemes. For example, the relation CHPA indicates a change from a property (such as ‘deep’) into an abstract object (such as ‘depth’). The relations shown in Figure 1 are association (ASSOC), attachment of object to object (ATTCH), property relationship (PROP), predicative concept specifying a plurality (PRED), experiencer (EXP), an informational process or object (MCONT), carrier of a state (SCAR), state specifier (SSPE), conceptual subordination for objects (SUB), conceptual subordination for situations (SUBS), neutral object (OBJ), temporal restriction for a situation (TEMP), and a function for the introduction of alternatives (\*ALTN1).

## 1.2 The WOCADI Parser

A syntactico-semantic parser is applied when preprocessing NL queries and for parsing all documents in the concept indexing approach and the network matching approach described in Section 2.1. The system uses the WOCADI parser (WOrd CLASS based DISambiguating parser; see for example Helbig and Hartrumpf (1997); Hartrumpf (2003)), which is based on the principles of Word Class Functional Analysis (WCFA). The parser generates for a given German sentence its semantic representation as a semantic network of the MultiNet formalism.

The NL analysis is supported by HaGenLex (Hartrumpf et al., 2003), a domain-independent computer lexicon linked to and supplemented by external sources of lexical and morphological information, in particular CELEX (Baayen et al., 1995) and GermaNet (Kunze and Wagner, 2001). HaGenLex includes:

- A lexicon with full morpho-syntactic and semantic information of more than 22,000 lexemes.
- A shallow lexicon containing words with morpho-syntactic information only. This lexicon comprises about 50,000 entries.
- Several lexicons with more than 200,000 proper nouns (including names of products, companies, countries, cities, etc.)

MultiNet (and therefore also HaGenLex) differentiates between homographs, polysemes, and meaning molecules.<sup>2</sup> The WOCADI parser provides powerful disambiguation modules (Hartrumpf, 2003, 2001), whose rules and statistics use syntactic and semantic information to disambiguate lexemes and structures.

<sup>2</sup>A meaning molecule is a regular polyseme with different meaning facets which can occur in the same sentence. For instance, two facets of ‘bank’ (building and legal person) are referred to in the sentence ‘The bank across the street charges a nominal fee for account management.’

The semantic network in Figure 1 illustrates several features of the parser: the disambiguation of a verb (the correct reading represented by the concept *berichten.2.2*<sup>3</sup>), the representation of a nominal compound *prüfungskandidat.1.1*<sup>4</sup> not contained in the lexicons together with its constituents *prüfung.1.1* and *kandidat.1.1*, and correct attachment and coordination of noun phrases. These features are important for our approach to translate queries: linguistic challenges for translation such as lexical, syntactic, or semantic ambiguities are already resolved by the parser.

### 1.3 The Database Independent Query Representation

To support access to a wide range of different target databases with different protocols and formal retrieval languages, the semantic network representation of a user query is transformed into an intermediate representation, a Database Independent Query Representation (DIQR). A DIQR expression comprises features typical for database queries:

- Attributes (so-called semantic access points), such as *author*, *publisher*, *title*, or *date-of-publication*.
- Term relations specifying how to search and match query terms. For example, the term relation '<' indicates that a matching document must contain a term with a value less than the given search term.
- Term types indicating a data type for a search term. Typical examples for term types are number, date, name, word, or phrase.
- Search terms identifying what terms a document representation should contain. Search terms include concepts (for example, *prüfung.1.1* / 'exam') and word forms (for example, "*Prüfungen*" / 'exams').
- Boolean operators in prefix notation for the combination of attributes, term relations, term types, search terms, or expressions to construct more complex expressions, for example 'AND' (conjunction) and 'OR' (disjunction). By convention, the operator 'AND' can be omitted, because it is assumed as a default.
- Optional numeric weights associated with search terms. These weights are used in information retrieval (IR) tasks to indicate how important a search term is considered in a query.

The DIQR is the result of a rule-based transformation of the semantic network representation using a RETE-based compiler and interpreter (the implementation is described in more detail by Leveling and Helbig (2002)). It is mapped to a query in a formal language the database management software supports (such as a query for the Z39.50 protocol, an SQL query, or a SOAP request), which is then submitted to the target system. For example, the semantic network in Figure 1 is transformed into the DIQR

```
((OR title abstract) = (AND (OR (phrase "psychologisch.1.1" "problem.1.1")
                                (word "stress.1.1")))
                       (OR (word "prüfungskandidat.1.1")
                           (wordlist "prüfung.1.1" "kandidat.1.1")
                           (word "prüfing.1.1"))) )
```

After expanding query terms with a disjunction of semantically related terms, the DIQR is normalized into a disjunctive normal form (DNF) and its components, written as conjunctions, are interpreted as query variants. The example above results in twelve disjunction-free query variants after normalization.

## 2 Monolingual GIRT Experiments (German – German)

### 2.1 Investigated Approaches

In the CLEF 2004 experiments, our focus is on comparing different indexing and matching techniques on different levels of abstraction for a document representation. Three different methods for indexing the GIRT document collection are employed:

<sup>3</sup>A lemma followed by a numerical suffix consisting of a numerical homograph identifier and a numerical polyseme identifier forms a so-called concept identifier (or concept ID) in HaGenLex.

<sup>4</sup>Compounds are written as one word in German.

1. **Indexing full word forms:** One database containing the German document collection (database GIRT4DE) and one containing the English document collection (database GIRT4EN) are created by indexing word forms from the documents. No document preprocessing takes place, i.e., no stemming, decomposition of compounds, or removal of stopwords.
2. **Indexing concepts:** The WOCADI parser produces semantic networks for the sentences in the title and abstract fields of the documents. From these semantic networks, concepts<sup>5</sup> are extracted and indexed (database GIRT4RDG). For compounds we add the concepts of their constituents (as determined by the compound module of the parser) to the index, e.g., we index *prüfungskandidat.1.1* in addition to *prüfung.1.1* and *kandidat.1.1*. To account for possible disambiguation errors, all word form readings determined by the morpholexical stage of the parser are chosen as index terms as well, but with a lower indexing weight. If the parser cannot construct a semantic network for a sentence, the latter terms are the only index terms to be added.
3. **Indexing semantic networks:** The parser returns the semantic network representations for a document's title and all sentences from its abstract (database InSicht). To reduce time and space requirements of this approach, each MultiNet (in its linearized or textual form) is simplified by omitting some semantic details less relevant for this application, and instance variables are replaced by constructed instance constants. Finally, to speed up matching even more by allowing optimized subset tests, every MultiNet is normalized by ensuring a canonical order of MultiNet terms. The resulting nets are indexed on the contained concepts to reduce the actual matching between the simplified networks of the documents and the query and thereby achieve acceptable answer times.

Note that for the bilingual experiments, only the first method can be applied, because currently WOCADI is restricted to analyzing texts in German.

## 2.2 Experimental Setup

The WOCADI parser and background knowledge represented as a single, large MultiNet allow to look up search terms semantically related to a given search term. Search term variants include orthographic variants (such as “*Schiffahrt*” and “*Schiffahrt*”), morphologic variants (for example, “*Stadt*” / ‘city’ and “*Städte*” / ‘cities’), and lexical variants (such as the synonyms *ansehen.2.3* / ‘lookup upon as’ and *betrachten.1.2* / ‘regard as’). The semantic similarity between two terms  $x$  and  $y$  is determined by their MultiNet relation  $rel$  and can be expressed as follows:

$$sim(x,y) = \begin{cases} 1.0 & \text{: if } x \text{ and } y \text{ are identical:} & x \text{ rel } y \text{ and } rel \in \{EQU, \dots\} \\ 0.95 & \text{: if } x \text{ is a synonym of } y: & x \text{ rel } y \text{ and } rel \in \{SYNO, \dots\} \\ 0.7 & \text{: if } x \text{ is a narrower term than } y: & x \text{ rel } y \text{ and } rel \in \{SUB, SUBS, PARS, \dots\} \\ 0.6 & \text{: if } x \text{ and } y \text{ are morphologically derived:} & x \text{ rel } y \text{ and } rel \in \{CHPA, CHPE, \dots\} \\ 0.5 & \text{: if } x \text{ is a broader term than } y: & y \text{ rel } x \text{ and } rel \in \{SUB, SUBS, PARS, \dots\} \\ 0.35 & \text{: if } y \text{ is a term otherwise related to } x: & x \text{ rel } y \text{ and } rel \in \{ASSOC, \dots\} \end{cases}$$

For concepts connected via a path of relations, the semantic similarity is calculated as the product of similarities along the path of relations connecting them.

Using the DIQR for the original query (OQ) as a starting point, the following steps are carried out as an automated retrieval strategy:

1. For each search term in OQ, the set of linguistically related concepts is obtained and OQ is expanded with the disjunction of search term variants (optionally weighted by semantic similarity). For a query translation, the translations of all term variants (concepts and words) are combined to produce a set of semantically related translations which serve to expand a query term in OQ (as will be described in Section 3). In this case, semantic similarities are replaced by translation scores to weight query terms.
2. The expanded OQ is normalized into DNF and its components, written as conjunctions, are interpreted as query variants. The query variants are ranked by their score (the semantic similarity

<sup>5</sup>A concept in MultiNet corresponds to one reading or lexeme in HaGenLex.

**Table 1:** Overview of parameter settings and results for monolingual GIRT experiments with the German document collection. The results displayed are the total number of documents retrieved and the mean average precision (MAP).

Run Identifier	Database Name	Parameters				# docs retrieved	MAP
FUHds1	GIRT4DE	Q-S	I-W	R-T	D-1	25000	0.2446
FUHdw1	GIRT4DE	Q-M	I-W	R-T	D-1	25000	0.2482
FUHdw2	GIRT4DE	Q-M	I-W	R-T	D-2	25000	0.2276
FUHdrw	GIRT4RDG	Q-M	I-C	R-E	D-1	25000	0.1162
FUHdm	InSicht	Q-M	I-N	R-E	D-3	1309	0.1126

between a query variant and OQ), which is computed as the product of the semantic similarities of their search terms, normalized by query length.

3. To construct a single database query, all search terms in the top ranked 250 query variants are collected in a word list to build an extended query. The documents found are retrieved until the result set exceeds a fixed size (here: 1000 documents). To perform multiple queries, the 250 top ranked query variants are separately used for retrieval. Documents scoring higher than the minimum score in the result set are retrieved and inserted into the result set.
4. Document scores are computed as the weighted sum of the database score ( $d_{score}$ ) (a standard *tf-idf* score as determined by the database ranking schema) and a query score ( $q_{score}$ ) (the semantic similarity between a query variant and OQ) for the current query variant:

$$doc_{score} = d_{score} \cdot d_w + q_{score} \cdot q_w$$

If multiple instances of a document are found and retrieved for different query variants, the maximum of the scores is taken.

## 2.3 Results

For the GIRT task in 2004, five experimental runs for the monolingual GIRT task were submitted. The experiments vary in the following parameter settings:

- a single query is created from all query variants (Q-S), or multiple queries (the query variants) are processed separately and their results are merged successively (Q-M)
- search terms and index terms are words (I-W), HaGenLex concepts (i.e., the numerical suffix is not stripped off the lexeme) (I-C), or concepts and relations from semantic networks (I-N)
- an exact search for search terms is performed (no truncation) (R-E), or a search for words beginning with the specified search term is performed (R-T) i.e., we use a so-called right truncation or prefix match
- the document score ( $doc_{score}$ ) is calculated as a weighted sum of database document score and query score with  $d_w = 0.7$  and  $q_w = 0.3$ . The query score is not normalized (D-1) or it is normalized by the query length (dividing  $q_{score}$  by the number of query terms) (D-2). A third variant uses  $d_w = 0$  and  $q_w = 1$  (D-3) to compute  $doc_{score}$ .

Table 1 gives an overview of the experiments performed for the German document collection with their results.

## 2.4 Brief Success and Failure Analysis

The following observations can be made from the retrieval results:

- The methods under investigation perform best in the order of word form indexing (FUHds1, FUHdw1, FUHdw2), concept indexing (FUHdwr), and semantic network matching (FUHdm) with respect to mean average precision (MAP). There is a general low performance of the experiments using disambiguated concepts (FUHdwr) and indexing and matching semantic networks (FUHdm).
- Experiments with multiple queries and with a single query have a similar performance (FUHds1 vs. FUHdw1). (There is not as much difference as in the experiments in 2003.)
- Normalization of the query score did not improve performance (FUHdw1 vs. FUHdw2).

Experiments with indexing concepts and matching semantic networks (FUHdwr and FUHdm) were expected to show a higher precision due to disambiguation of concepts in both queries and documents. A plausible explanation for the observed results is that these experiments rely on WOCADI for analysis of both queries and documents. We have observed that documents in the GIRT collection (German titles and abstracts) are difficult to parse or to represent as semantic networks, because their abstracts often contain grammatically incorrect, malformed language (for instance, the table of contents of a book) or spelling errors. For example, of 60,702 words occurring in the GIRT documents with a frequency of one, we judged 11,589 as spelling errors. The remaining words mainly are words unknown to HaGenLex, foreign words, and proper nouns not in the name lexicons. Due to time constraints, we did not investigate words with a higher frequency.

WOCADI produced full semantic networks for 34.3% of the 1,111,121 sentences (with 23,088,562 words) in the GIRT documents, partial semantic networks<sup>6</sup> for 23.9% of the sentences, and no semantic networks for the remaining 41.8% of the sentences. These results are significantly worse than for other corpora (for example see Hartrumpf (2004) for WOCADI's parse statistics on the QA@CLEF 2004 newspaper and newswire corpora). One promising extension would be to include the available partial semantic networks in a modified matching procedure.

The MultiNet indexing experiment is based upon matching the semantic network for a query with semantic networks for a document on a per-sentence basis, i.e., one semantic network per sentence in a query or document is matched. However, relevant documents can contain search terms not co-occurring in the same sentence, which currently will not be found. Queries may consist of multiple sentences. For example, a topic's narrative typically contains several sentences and can be considered a query. The problems of the preceding paragraph can be solved: Semantic networks are not restricted to represent the meaning of one sentence and the WOCADI parser is capable of analyzing a text consisting of multiple sentences (including coreference resolution) and returning a single semantic network. So, there are several directions for improving the MultiNet matching approach.

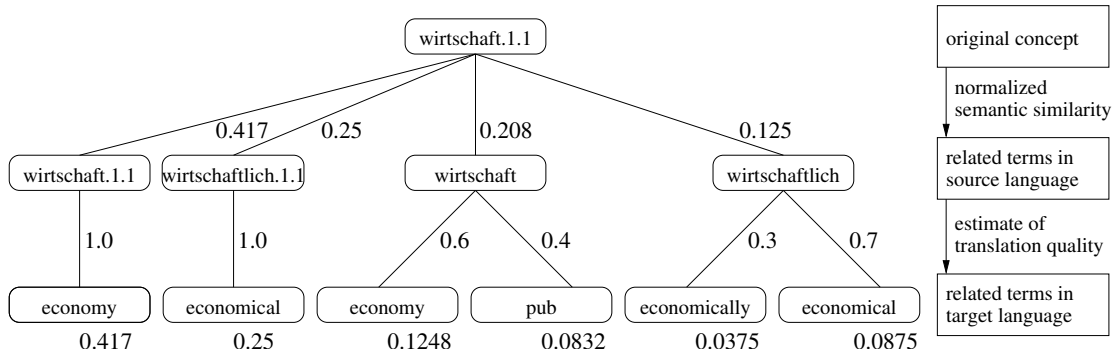
To summarize, indexing concepts or indexing and matching semantic networks did not show a higher precision than the traditional IR approaches. We do not draw the conclusion that these approaches will not perform better in the near future (as they still have more potential for improvements) or that they are not suited for information retrieval. It remains to be seen if this behavior is specific to the GIRT document corpus. Indexing full text, such as newspaper articles may provide a better basis for experiments with matching and indexing semantic networks.

### 3 Bilingual GIRT Experiments (German – English)

For the bilingual retrieval experiments with the GIRT document collection, we apply a dictionary-based translation of the concepts in the DIQR. Currently there is no English version of HaGenLex, but there is an incomplete mapping between HaGenLex concepts and GermaNet concepts (Osswald, 2004). GermaNet is the German part of EuroWordNet (Vossen, 1998). This translation lexicon contains about 10,000 translations of HaGenLex concepts into EuroWordNet concepts.<sup>7</sup> The high quality concept translation lexicon was combined with a translation word list with about 110,000 entries compiled from several resources for translating German word forms into English (from LEO: <http://dict.leo.org>; DICT resources: <http://dict.tu-chemnitz.de>). With these resources, a concept translation lexicon and a translation word list, there are two ways to translate a HaGenLex lexeme:

<sup>6</sup>WOCADI tries to produce such networks in a special chunk mode (or shallow parsing mode) when a full parse has failed.

<sup>7</sup>With this mapping between HaGenLex and EuroWordNet concepts, disambiguation information in the target language will be still available. If standard machine translation software were to be used, readings would not be differentiated or the differentiation would differ from the concepts and meanings used in HaGenLex.



**Figure 2:** Tree representation for translating the concept *wirtschaft.1.1*. A translation score for a concept is computed as the sum of products of edge markings on the path to the leaf nodes containing this concept. The resulting order of possible translations is ‘*economy*’ (score:  $0.417 \cdot 1.0 + 0.208 \cdot 0.6 = 0.5418$ ), ‘*economical*’ (0.3375), ‘*pub*’ (0.0832), and ‘*economically*’ (0.0375). Cross-part-of-speech translations are included, but disambiguation information may be partly lost. For instance, the English word ‘*pub*’ is actually a translation of the HaGenLex concept *wirtschaft.1.2*. Some subbranches have been pruned from this example tree for readability (e.g., the branches with the semantically related concepts *ökonomie.1.1* and *ökonomisch.1.1* have been omitted).

1. Remove the numerical suffix from the HaGenLex lexeme and try to find a translation in the word translations. For example, among the translations for the word “*arbeit*” are the English words ‘*work*’, ‘*job*’, ‘*occupation*’, and ‘*exam*’.
2. Look up the EuroWordNet concept correspondences for the HaGenLex lexeme. The translation for the concept *arbeit.1.1*, for instance, includes the mapping to the correct EuroWordNet concepts for ‘*work*’ and ‘*labor*’.

We combine both methods and create a tree representation to find a set of semantically related words in the target language for a given concept in the source language. The root of the tree denotes the concept for which semantically related translations are to be found. All immediate successors of the root node represent concepts and base forms (concepts without the numerical suffix) that are semantically related to the root concept. For words the semantic similarity is estimated to be half the semantic similarity between original concept and the concept corresponding to the word. The corresponding arcs are associated with a numeric value subject to a probabilistic or frequentistic interpretation (i.e., the semantic similarity normalized to the interval  $[0, 1]$ ). Leaves are successors of these nodes and represent the translations found by applying one of the two methods mentioned above. Their arcs are marked with either a normalized frequency or an estimate of the translation quality dependent of the source of the translation. Leaf nodes are marked with the product of numerical values on the edges from the root to the leaf node.

A ranking of all semantically related translations is obtained by collecting all leaf concepts and their translation score is calculated as the sum of all leaf values.<sup>8</sup> For example, the ranking of translations for the concept *wirtschaft.1.1* by their translation scores is ‘*economy*’, ‘*economical*’, ‘*pub*’, and ‘*economically*’ (see Figure 2). Extending this approach, leaf nodes in the target language might be expanded by semantically related concepts as well (e.g., a leaf can be expanded by concepts in its EuroWordNet synset). For the bilingual experiments, translation scores replace semantic similarities in the retrieval strategy described in Section 2.2.

### 3.1 Experimental Setup

Our first participation in the Bilingual GIRT task (matching German topics against the English data) consists of trying various parameter settings for translating German query concepts into English. The English

<sup>8</sup>Assuming conditional independence between the nodes, the tree can be interpreted as a decision tree, in which case computing the translation score is equivalent to applying Bayes’ theorem for computing probabilities.

**Table 2:** Overview of parameter settings and results for bilingual GIRT experiments with German queries and the English document collection.

Run Identifier	Database Name	Parameters				# docs retrieved	MAP
FUHe1	GIRT4EN	G-5	E-5	W-U	Q-M	18006	0.0351
FUHe2	GIRT4EN	G-all	E-5	W-T	Q-M	18252	0.0369
FUHe3	GIRT4EN	G-5	E-all	W-T	Q-M	18911	0.0365
FUHe4	GIRT4EN	G-5	E-5	W-T	Q-M	18735	0.0366
FUHe5	GIRT4EN	G-5	E-5	W-T	Q-S	12766	0.0229
FUHe1 re-run	GIRT4EN	G-5	E-5	W-U	Q-M	23805	0.1117
FUHe2 re-run	GIRT4EN	G-all	E-5	W-T	Q-M	23805	0.1135
FUHe3 re-run	GIRT4EN	G-5	E-all	W-T	Q-M	23805	0.1288
FUHe4 re-run	GIRT4EN	G-5	E-5	W-T	Q-M	23805	0.1275
FUHe5 re-run	GIRT4EN	G-5	E-5	W-T	Q-S	23805	0.1104

GIRT document collection was indexed as is, i.e., English words were indexed (I-W). Document scores are calculated as in the monolingual experiments (D-2) and truncation is applied to the search terms (R-T).

The experimental parameters varied are:

- a single query is created by combining all query variants (Q-S), or multiple query variants (Q-M) are processed
- for a search term, all semantically related terms are used as query term variants (G-all), or the top five semantically related terms (ranked by semantic similarity) are used (G-5)
- all translations found are used as query term translations in a query (E-all), or the best five translations are used (E-5)
- the translation scores are used to weight query search terms (W-T), or query terms are not weighted (W-U)

## 3.2 Results

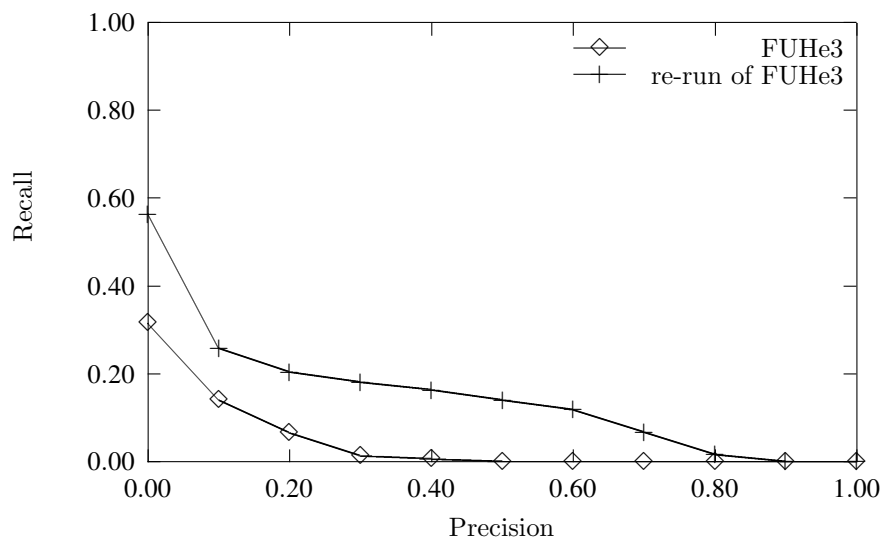
Five runs for the bilingual GIRT task (German – English) were submitted for relevance assessment. The results of our official experiments in the bilingual GIRT task are seemingly discouraging. During an analysis of possible causes of failures it became obvious that due to a software problem with the freely available database indexing tool we employed, the database index for the database GIRT4EN was not constructed correctly. All SGML data files starting with “LI4E” were not indexed (which means that only 15,955 documents out of 151,319 were indexed at all). To obtain meaningful results, we started re-runs of our official experiments after fixing the software problems. Parameter settings and results for the official experiments and the re-runs are shown in Table 2 and in Figure 3, respectively.

## 3.3 Brief Failure Analysis

This brief failure analysis refers to our unofficial re-runs for the bilingual task.

- There is a lower performance for the bilingual experiments compared to the monolingual experiments.
- Queries that could not be processed successfully in German, could not be processed in English as well (for example, topic 115).
- For one query, no relevant documents were found in all experiments (topic 112). For several queries, less than 1000 documents were retrieved.





**Figure 3:** Recall-precision graph for one official bilingual GIRT experiment (German-English) and its re-run. Because results of runs (re-runs) are not significantly different, only one run (re-run) is shown.

The bilingual experiments employ a dictionary based per-word or per-concept translation. Search failures for some topics indicate that there are too few translations available. For example, the same number of documents was retrieved in all bilingual experiments, because less than 1000 or no documents were retrieved for the same topics in all experiments. The missing translations are one major reason why the performance in comparison with the monolingual experiments is lower. The experiments should be repeated when the translation of concepts into EuroWordNet concepts has been completed with respect to HaGenLex coverage.

Translation ambiguities lead to noise in the results, which could be reduced if certain syntactic or semantic structures are treated differently for a translation. In particular, German noun compounds and adjective-noun phrases were translated by translating their constituents, but should be treated depending on their semantic context. Consider, for example the compounds and their (correct) translations “*Klimaänderung*” / ‘*climate change*’, “*Klimaanlage*” / ‘*air conditioning (system)*’, “*Klimakammer*” / ‘*climatic chamber*’, ‘*environmental chamber*’; or “*Sonnenenergie*” / ‘*solar energy*’, “*Sonnenbrand*” / ‘*sunburn*’, “*Sonnencreme*” / ‘*suntan creme*’, ‘*sunscreen*’. In these compounds, the common German constituent is translated differently, although its meaning should be represented by the same concept. Simply adding all translation alternatives for a compound constituent to expand a query adds too much noise to the results. In a similar form, this problem arises for adjective-noun phrases.

## 4 Conclusion

In comparison with the results for the monolingual GIRT task in 2003, performance with respect to the best MAP has improved (0.2482 in 2004 vs. 0.2064 in 2003 for a similar run). Monolingual experiments using word-based retrieval (i.e., retrieval based on indexed full-forms) have a higher MAP than both the experiment with concept indexing and indexing of semantic networks. However, a comparison with other corpora suggests that the low performance of indexing semantic networks might be specific to the GIRT document collection. In contrast to the traditional approach we tested, the semantic network approach aims at representing the meaning of a document. For this approach, there are obvious improvements for further experiments, including matching across several sentences (with coreference resolution for multiple sentences) and matching partial semantic networks.

The re-runs of the official bilingual experiments showed encouraging results. After completing the mapping of HaGenLex concepts to readings of EuroWordNet, other languages will be available for cross-language information retrieval experiments with our system. Additional lexeme translations and improved

methods to translate multi-word expressions and compounds will be integrated to increase performance for bilingual IR.

## References

- Baayen, R. Harald; Richard Piepenbrock; and Leon Gulikers (1995). *The CELEX Lexical Database. Release 2 (CD-ROM)*. Philadelphia, Pennsylvania: Linguistic Data Consortium, University of Pennsylvania.
- Hartrumpf, Sven (2001). Coreference resolution with syntactico-semantic rules and corpus statistics. In *Proceedings of the Fifth Computational Natural Language Learning Workshop (CoNLL-2001)*, pp. 137–144. Toulouse, France.
- Hartrumpf, Sven (2003). *Hybrid Disambiguation in Natural Language Analysis*. Osnabrück, Germany: Der Andere Verlag.
- Hartrumpf, Sven (2004). Question answering using sentence parsing and semantic network matching. In *Results of the CLEF 2004 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2004 Workshop* (edited by Peters, Carol). Bath, England.
- Hartrumpf, Sven; Hermann Helbig; and Rainer Osswald (2003). The semantically based computer lexicon HaGenLex – Structure and technological environment. *Traitement automatique des langues*, 44(2):81–105.
- Helbig, Hermann (2001). *Die semantische Struktur natürlicher Sprache: Wissensrepräsentation mit Multi-Net*. Berlin: Springer.
- Helbig, Hermann and Carsten Gnörlich (2002). Multilayered extended semantic networks as a language for meaning representation in NLP systems. In *Computational Linguistics and Intelligent Text Processing (CICLing 2002)* (edited by Gelbukh, Alexander), volume 2276 of LNCS, pp. 69–85. Berlin: Springer.
- Helbig, Hermann and Sven Hartrumpf (1997). Word class functions for syntactic-semantic analysis. In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP'97)*, pp. 312–317. Tzigov Chark, Bulgaria.
- Kluck, Michael and Fredric C. Gey (2001). The domain-specific task of CLEF — specific evaluation strategies in cross-language information retrieval. In *Cross-Language Information Retrieval and Evaluation. Workshop of the Cross-Language Information Evaluation Forum (CLEF 2000)* (edited by Peters, Carol), volume 2069 of LNCS, pp. 48–56. Berlin: Springer.
- Kunze, Claudia and Andreas Wagner (2001). Anwendungsperspektiven des GermaNet, eines lexikalisch-semantischen Netzes für das Deutsche. In *Chancen und Perspektiven computergestützter Lexikographie* (edited by Lemberg, Ingrid; Bernhard Schröder; and Angelika Storrer), volume 107 of *Lexicographica Series Maior*, pp. 229–246. Tübingen, Germany: Niemeyer.
- Leveling, Johannes (2003). University of Hagen at CLEF 2003: Natural language access to the GIRT4 data. In *Results of the CLEF 2003 Cross-Language System Evaluation Campaign, Working Notes for the CLEF 2003 Workshop* (edited by Peters, Carol), pp. 253–262. Trondheim, Norway.
- Leveling, Johannes and Hermann Helbig (2002). A robust natural language interface for access to bibliographic databases. In *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002)* (edited by Callaos, Nagib; Maurice Margenstern; and Belkis Sanchez), volume XI, pp. 133–138. International Institute of Informatics and Systemics (IIS), Orlando, Florida.
- Osswald, Rainer (2004). Die Verwendung von GermaNet zur Pflege und Erweiterung des Computerlexikons HaGenLex. *LDV Forum*, 19(1):43–51.
- Vossen, Piek (editor) (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht, The Netherlands: Kluwer.