# Why Information Retrieval Needs Cognitive Science: A Call to Arms.

**Eduard Hoenkamp (hoenkamp@acm.org)**
Nijmegen Institute for Cognition and Information (NICI); Montessorilaan 3
6525 HR Nijmegen, the Netherlands

## Abstract

Much of today's success in Information Retrieval (IR) comes from a hard approach: employing blazingly fast machines, ever more refined statistics, and increasingly powerful classification schemes. In recent years, however, the hard approach has entered a phase of diminishing returns.
This paper explores a softer alternative which, we argue, is still in the phase of increasing returns. As the quality of an IR system is ultimately decided by its users, the approach starts from how these users structure information. Interestingly, for this approach many useful principles are readily available in the psychological literature.
We illustrate the approach with three examples. The first applies the cognitive status of 'complex nominals' to improve search results by automatically constructing specialized queries. The second shows how the connection between language and imagery at the 'basic level' can be used for multimedia retrieval on the World Wide Web. The final example employs the notion of 'semantic space' to make retrieval more effective especially for large scale corpora. In each example the results were substantial. The cases we studied illustrate how an approach to information retrieval based on cognitive principles can lead to significant, immediate, and fundamental results. It shows how prolific the application of cognitive science to the core of IR can be, and we believe that both disciplines stand to benefit from this approach.

**Keywords:** Information retrieval; cognitive approach; increasing returns; case studies.

## Introduction

Browsing through proceedings of the major conferences on information retrieval[1] one is confronted with a hard approach: ever faster machinery, increasingly refined statistical techniques, algorithms tweaked to specific domains, applied linear algebra, and more and more powerful classification methods. At the same time, one cannot escape the impression that more and more effort is needed to yield relatively small results, or in economic terms, diminishing returns. This could mean that the IR problem is practically solved, or maybe that the field is ready for a paradigm shift away from the dominant hard approach. The former will raise the eyebrows of anyone who looks up information on the Internet. The latter conclusion is perhaps premature. Whichever is the case, the alternative route is gaining attention as worthwhile to pursue.
That cognitive science has an influence on IR is clear: User evaluation studies and studies in interface design are legion. But note that such studies are more an application of the experimental paradigm than an innovative approach to improve the core of IR. Also the influence of IR on cognitive science is evident: Most notably, a decade of latent semantic analysis (LSA) shows how the data-driven approach within IR has inspired studies of human cognition (see (Dumais, 2003) for a comprehensive overview). In this paper we want to show an approach different from these two, one where the use of known cognitive principles can lead to more effective information retrieval. We think such research is direly needed to complement the hard approach, but the examples we found are very few. A notable exception is the theory of a 'cognitive space' (Ingwersen, 1999; Gärdenfors, 2000) as alternative to the document space model of IR (Salton, 1988). It is an example where an old psychological theory, namely Osgood's (1957) semantic differential, is elaborated into an innovative application for IR. It also illustrates the main point of the present paper: no new cognitive theories need to be developed to innovate IR, as there are enough cognitive principles readily available in the literature.

The sections that follow are self-contained presentations for each case study. However, for a complete coverage of the three cases, the space limitations were too tight. As these details can be found in the work we published elsewhere, we think we can argue our main point: how the straightforward application of cognitive science can lead to new insights and innovation in information retrieval.

## The case of Complex Nominals

Intuitively, words forming noun phrases (NP's) are a richer and more precise representation of meaning than keywords: 'horse' and 'race' may be related, but 'horse race' and 'race horse' carry more circumscribed meanings than the words in isolation. Hence, several researchers have suggested that using noun phrases instead of keywords, may improve retrieval effectiveness. Yet, empirical studies using noun phrases as queries did not confirm this intuition (Croft, 1995).

[1] Notably the SIGIR Conference on Research and Development in Information Retrieval, the TExt Retrieval Conferences (TREC), and the Digital Libraries Conference.

The research we will present here isolates a general, ubiquitous, and productive sub-category of noun phrases for which we found the intuition to hold: the *complex nominal (CN)*. Complex Nominals, such as nominalizations and noun compounds have been studied for centuries. What makes these constituents so special?

*A. Findings from linguistics:*
- *CN*s behave grammatically as single nouns,
- occur (probably) universally over languages,
- evoke a preferred reading,
- are correctly used in the basic variety[2].

*B. Findings from Cognitive Science:*
- *CN*s are efficient in communication,
- are very productive,
- can act as nonce words (used only once).

*C. Findings from Information Filtering*
- *CN*s can be generated from a semantic representation,
- reduce user interaction,
- attain a precision[3] noun phrases generally don't.

The latter findings come from our information filtering project PROFILE (Hoenkamp, Schomaker, van Bommel, Koster, & van der Weide, 1996).

The rationale behind using noun phrases as queries is not only that they are more precise than the keywords in isolation. Another reason is that people use NP's all the time to identify referents in the information they want to convey. Since PROFILE (or any information filter) is meant to find information on behalf of the user, the filter is on its own to construct the query. Hence, it may have to express concepts for which is does not have a vocabulary. This is exactly the situation where people use complex nominals pervasively. But how to construct interpretable queries without human intervention?

## Generating Complex Nominals from a conceptual representation

Several researchers (Liddy & Myaeng, 1993; Gardiner, Riedl, & Slagle, 1994) have described systems that translate conceptual graphs into a query to search the TREC corpus (using boolean search with proximity). PROFILE instead uses the CYC formalism (Lenat, 1995) as conceptual representation to generate *CN*s. Much is known about the way people process novel *CN*'s, see e.g. (Ryder, 1994) and (Adams, 1973) from which we selected our *CN*'s. Our mechanism for generating *CN*'s from a conceptual representation is based on Levy's (1978) theory about complex nominals. It is a transformational grammar enriched with so-called 'recoverable deletable predicates' (RDP), and a set of meaning preserving transformations. The algorithm proceeds in three steps (Hoenkamp & de Groot, 2000): (1) Objects in the CYC representation are lexicalized via WORDNET (Miller, 1995), (2) each CYC relation is mapped to an RDP, which together with the previous step produces a deep structure, (3) the permitted transformations are applied to the deep structure to yield all the *CN*'s that can express the conceptual representation.

## Proximity vs. Coherence

Search engines often provide operators to confine the matching of keywords to passages where the keywords are within a certain proximity, e.g. ALTAVISTA's NEAR operator. We will compare this *proximity matching* to what we will call *coherence matching*. *Proximity Matching* with $N_1N_2$ is like ALTAVISTA's "$N_1$ NEAR $N_2$". For example, "picture NEAR book" matches "She bought a picture book of Italian Art", but also "He had his picture taken carrying a book". *Coherence Matching* with $N_1N_2$ occurs if a passage contains one of: (1) an **exact** match with $N_1N_2$, or (2) a match with a **syntactic variation**, e.g. "picture book" matches "book with pictures" as in "She had chosen the book with the rare pictures" (3) a match which is a **semantic variation**, e.g. "picture book" matches "volume of portraits" as in "He published a volume of excellent portraits of contemporary scientists". (Examples from actual retrieval results.) The *syntactic variations* are formed through linguistic transformations applied to the deep structure. The *semantic variations* emerge through the variety of words in WORDNET that express the same word sense.

## Finding relevant passages

In the example above, proximity matching with "picture book" located "He had his picture taken carrying a book". Coherence matching removed the irrelevant result for the reasons we discussed. The conjecture, then, is that *coherence matching improves precision.*

To compare the two techniques, we had to select a document collection that would promote reliable and valid results. An ideal collection would be: (1) a corpus that is grammatically tagged, making the experiment independent of the power and correctness of a parser, (2) which contains a substantial portion of spoken language, so that novel *CN*'s are likely to occur, (3) which contains a variety of subject areas, to insure generality of the results. The British National Corpus (BNC) is such a collection; a 100 million word collection of samples of written and spoken language from a wide range of sources. The whole collection has been grammatically tagged. To be even more accurate, we used the 'BNC sampler', a 10% subset of the corpus, where the grammatical tags have been manually checked, and which consists of about an equal amount of spoken and written texts.

Adams' (1973) extensive overview of word formation distinguishes thirteen classes of *CN*'s on the basis of the semantic relationship between the constituent nouns. We took an example from each class, and we produced an exhaustive set of syntactic and semantic variations.

Next we simulated an existing WWW search engine: We first had the BNC sampler indexed by AltaVista. Then we reverse engineered our proximity matcher until

[2] The 'basic variety' is a well-structured, remarkably efficient and simple form of language, that adult second language learners universally develop, when not under the influence of a particular teaching method (Klein & Perdue, 1997)

[3] The dominant evaluation metrics in IR are *precision* and *recall*. When documents are retrieved, precision is the proportion of the retrieved documents that are actually relevant, and recall is the proportion of the relevant documents that are actually retrieved.

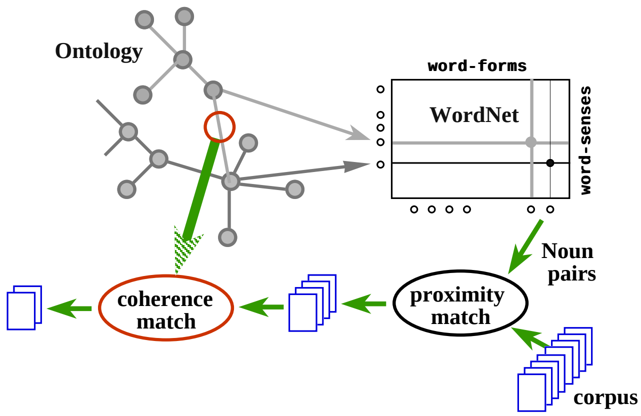


Figure 1: Coherence matching for a compound concept containing two sub-concepts and a relation between them. Sub-concepts are translated via WORDNET to pairs of keywords, thus losing their original relationship. Proximity matching locates potentially relevant passages in the corpus. Coherence matching subsequently reinstates the original relationships (dashed arrow), removing passages where the relationship does not hold.

it produced the same hits. Using this proximity matcher, every *CN* exemplar we had constructed was matched against each document in the BNC sampler. All matching passages were manually scored by two raters, and marked as relevant if the raters agreed. Next precision was calculated. This was repeated for all sets of *CN*'s. The results are in the top row of Table **??**. In a second

| | relevant passages | irrelevant passages | average precision |
|---|---|---|---|
| proximity matching | 36 | 152 | .43 |
| coherence matching | 27 | 38 | .72 |

Table 1: Comparison of proximity and coherence matching for the 'BNC sampler'. Matching was performed with exhaustive semantic expansion, and limited stemming (-ing, -er, etc.). Coherence matching shows a precision gain from .43 to .72, which is significant (Wilcoxon, $N = 13$, one-tailed, $p < .01$).

pass, coherence matching was performed over the documents already found during proximity matching (see Figure **??**). The matching passages were scored by two raters, as before, and the results are summarized in the bottom row of Table **??**. The significance test: We don't know the distribution of relevant passage in the BNC sampler per query, so we need a non-parametric test. Further, coherence matching occurs over the results of proximity matching (as it is a second pass), so the observations are correlated. Hence the Wilcoxon test, a non-parametric test for correlated paired observations. The considerable increase in precision is significant at the level of $p < .01$.

Given this positive result, we stepped up the scale of the

| | relevant passages | irrelevant passages | average precision |
|---|---|---|---|
| proximity matching | 633 | 843 | .53 |
| coherence matching | 541 | 126 | .77 |

Table 2: Comparison of proximity and coherence matching for the complete BNC. Coherence matching again showed a significant precision gain over proximity matching (Wilcoxon, $N = 11$, one-tailed, $p < .01$).

corpus and repeated the experimental procedure for the complete BNC. The results are summarized in table **??**. Again a significant increase in precision ($p < .01$).

This concludes the first case that demonstrates our main point: that information access can be substantially improved by applying cognitive principles readily available in the literature. Onward to the next case.

## The case of the 'Basic Level Category'

When you hear the word "chair", you may imagine a chair. You can mention its parts, indicate its height, perhaps mime how you would sit down in it. Hear the word "kitchen chair" instead, and not much will change. But for the word "furniture" the situation changes radically: no simple image pops up, and there is little to mention in terms of parts or specific motor activity. The level in a hierarchy of concepts at which such sudden proliferation of attributes occurs was investigated by Rosch (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), who called this the *basic level*. Several of the initial claims about the basic level have been subject to debate (Tversky & Hemenway, 1984; Murphy & Smith, 1982). Uncontested, however, are Rosch's observations about the striking connections at the basic level between imagery and language, to wit:

*A. Findings from Imagery:*
• The basic level appears the most abstract level for which an image can represent a class as a whole (Peterson & Graham, 1974),
• When just the name of an object is mentioned to a subject, about the same attributes are listed as when that object is visually present (Rosch et al., 1976).
*B. Findings from Language studies:*
• Many more attributes are listed for words at the basic level than for the super-ordinate, and few additional for the subordinate (Rosch et al., 1976),
• For physical objects and organisms, *parts* notably proliferate at the basic level (Tversky & Hemenway, 1984),
• When people have to name a picture of an object at the subordinate level, they choose the word for the basic level (Rosch et al., 1976).

As keywords have become a serious limitation when searching non-textual material on the internet, we hypothesized that the 'basic level' might be a way to derive appropriate keywords for image retrieval.

## Content Retrieval from the Web via the basic level

In the context of content delivery (text as well as images), we formulated several conjectures that follow from the properties of the basic level (cf. the summary in the previous section):
• Documents about parts of a basic level category can be retrieved by searching for the basic level word,
• Images of a basic level category can be retrieved by searching for the basic level word,
• The previous two cases should show notably higher precision for unambiguous basic level words than for polysemous basic level words.

As domain for our study we chose the *aircraft* domain. First, its basic level words: airship, airplane, balloon, and helicopter are unambiguous, which is conducive to precision. Second, their underlying concepts can be clearly distinguished through two simple criteria (for example, 'flying because lighter than air, not dirigible' distinguishes balloon from the other three). Let us take a closer look at each of these conjectures.

**Conjecture about discovering parts of objects.** The first conjecture expresses the connection between words and images by combining (1) the evidence that objects are identified by recognizing their parts (Biederman, 1987), and (2) the evidence that parts are important in distinguishing the basic level per se (Tversky & Hemenway, 1984). To test the conjecture, we set out to find parts of an airplane (the concept) on WWW via the basic level word 'airplane'. We informally collected a layman's aircraft ontology, which contains words such as wing, tail, gear, rudder etc. The automated search proceeded as follows. A search engine was queried until about ten thousand documents were retrieved that gave a hit on 'airplane' (the basic level word). These documents were indexed, followed by dimension reduction (SVD). Then words that loaded most heavily on the principal axes were clustered. Words that loaded higher than a threshold were retained. The algorithm set this threshold such that the words expressing parts already in the ontology were retained. This way, sixty words remained. We inspected these words by hand and found that words indicating parts of an airplane were prominently present. Especially noteworthy is that several new parts were discovered that were not present in the original ontology, such as 'aileron' and 'elevator'. This confirms the first conjecture, namely that parts of an object can be found by searching for the basic-level word. In addition, it shows an example where a lacuna in the ontology was discovered by a fully automated procedure (Hoenkamp, 1998).

**Conjecture about retrieving images.** To investigate the second conjecture, we used the observation by Rosch et al. (1976) that the outlines of objects within the basic level look alike. In the PROFILE representation (our 'ontology') concepts can have two special pointers: (1) if a concept occurs as a word sense in WORDNET then it has a pointer to that word sense. This way, WORDNET was used to translate concepts into keywords. (2) If a concept is a basic level category, than it has a pointer to a stereotypical image. (People tend to choose a fairly standard viewpoint when illustrating material).

The image retrieval prototype proceeded in two stages: a *retrieval stage* and a *selection stage*. In the retrieval stage a user sketches the outline of some aircraft. The outline is compared to the stereotype images stored with the ontology. The concept corresponding to the best match is looked up, and its index into WORDNET is followed. The word(s) expressing the word sense are then sent to several search engines. In the selection stage the hits are retrieved, and the images are isolated (extensions .jpg, .gif etc). These images are compared to the outline the user had originally drawn, and the pictures that match are shown to the user for verification. In our experiments the precision of the retrieved images was low. However, in contrast to textual documents, this does not pose a problem to users: To find the relevant documents among e.g. 400 text documents is hard work, to spot the relevant pictures among 400 pictures (for example in a grid of 20 by 20) is easy.

**Conjecture about polysemy and precision.** The third conjecture we made predicts that the good results in the previous sections may not transfer in case of polysemy, as the precision is inherently much lower. We compared the aircraft domain with its unambiguous basic level words to a domain with highly polysemous basic level words, namely furniture. We compared balloon and helicopter to chair and table as keywords, and retrieved about 500 documents for each. We isolated the pictures from the documents, tallied the relevant pictures for each and calculated precision (see Table **??**). When we compared the aircraft domain to that of furniture, the average precision fell from about 20% to a meager 2%. This may look disastrous, but the basic level itself provides a remedy for this lack of precision: Since the basic level is the highest level for which an image can be construed, an image to illustrate the superordinate level must be at the basic level. So instead of the highly polysemous words 'chair' or 'table' we can use 'furniture' with its low polysemy to retrieve images at the lower level. Indeed, we found (Hoenkamp, Stegeman, & Schomaker, 1999) that the number of relevant documents markedly increased this way.

| **Basic level word** | **number of docs** | **number of images** | **precision (%)** |
|---|---|---|---|
| Balloon | 436 | 153 | 13 |
| Helicopter | 432 | 438 | 28 |
| Chair | 500 | 33 | 2 |
| Table | 500 | 7 | 1 |

Table 3: Precision decreases dramatically with increasing polysemy (aircraft vs furniture).

This section demonstrated how exploiting cognitive properties of the basic level, enabled us to automatically generate appropriate keywords for image retrieval. Even for the fastest search engines it would have been impossible to match our example images to the billions of documents on WWW. So also in the case of multimedia retrieval the cognitive approach demonstrated an improvement over the hard approach to IR.

## The case of 'Semantic Space'

Our everyday search engines are based on the so-called 'vector space model' (Salton, 1988), in which documents are seen as points in a metric space spanned by the terms (words) in the corpus. In recent years, probabilistic language modeling is gaining recognition as a viable alternative. In this model a document is seen as a sample from a word sequence generated according to some distribution. These distributions would assign different probabilities to a sequence of query terms. And for a given query, the documents can be ranked according to the probabilities their distributions would assign. So a language model for IR needs to describe (1) the document model, i.e. the distribution over terms, and (2) the query model, i.e. how a probability is assigned given the query terms. In the hard approach to IR, a variety of proposals have been published regarding the choice for (1) and (2). But we will look at these points from a cognitive standpoint.

### Deriving a query model

The strength of language models lie in their well-understood formalisms and mathematical rigor. Their weakness, however, lies in the additional assumptions required to make the formalisms tractable in large scale applications. Why some assumptions work well and others don't is not always clear. In our cognitive approach, we capitalize on two properties of the documents to be retrieved. One is that texts are samples from a natural language corpus, hence have surface constraints. The other is that texts represent content, hence have underlying semantic dependencies. To our surprise, we discovered that the language models proposed to date have overlooked either or both properties of the very material they are trying to model.

**The document model as an ergodic chain.** Many cognitive phenomena can be well understood in terms of word-pairs, cf. the research on memory (Shiffrin & Steyvers, 1998) and on the 'semantic space' (Burgess, Livesay, & Lund, 1998). If the probability of a term depends only on the preceding term, then one can define the distribution as a Markov chain with the terms as states. The subsequent states, then, are sequences of terms which form documents. Two observations about natural language seem so obvious that they are easily overlooked: (1) Words can be separated by any number of intermediate words. Therefore, the Markov chain producing the words is *aperiodic*, and (2) You can always get from one word to another by continuing to produce text. Consequently, the Markov chain is *irreducible.* A Markov chain with both properties is called *ergodic*, and it has the property that in the long run it reaches a stationary distribution, irrespective of the initial state. It follows from this observation that if we could know the transition probabilities of terms (in a document or corpus) we could compute the stationary distribution to use as query model. (And we don't need invent simplifying assumptions to approximate the distribution.)

**Where do the priors come from?** In our experiments the prior probabilities are derived from the 'Hyperspace Analog to Language' (HAL) representation for a corpus (Burgess et al., 1998). The representation is computed by sliding a window over the documents and assigning weights to word pairs, inversely to the distance from each word to every other in the window. We normalize the distances to use as transition probabilities for Markov chain. Note how the two special properties of texts have been encoded in the Markov chain: From the underlying semantic dependencies follow the transition probabilities, and from the surface structure follows the ergodicity.

### Validation on a large scale corpus

The experiment we did to validate the Markov approach is rather technical, and perhaps requires more than a passing familiarity with IR. Yet we would like to show how the cognitive science orientation stands up to a corpus the size that has so far only been approached by the hard branch of IR (about 50 million words). We conducted an experiment on relevance feedback with re-weighting. In this paradigm, a rather crude method is used to collect a set of so-called pseudo-relevant documents. Then the weights used for ranking the documents are re-weighted. We repeated a well-known TREC experiment except that we used our cognitively based calculation for re-weighting.

| AP89 topics | Baseline | Robertson's TREC-3 | Stationary kernel |
|---|---|---|---|
| 101-150 | 0.1806 | 0.2298 (+27%) | 0.2594 (+44%) |
| 151-200 | 0.2244 | 0.2386 (+6%) | 0.2726 (+22%) |

Table 4: Average precision for the two re-weighting schemes compared to the baseline

*Material*: TREC corpus AP89, and two sets of topics (101-150, 151-200).
*Procedure*: We applied our method to Robertson's (1992) query expansion approach in TREC-3 as follows. We took BM25 as baseline, and the 20 top-ranked documents as the relevant set. First, Robertson's re-weighting scheme for Okapi was applied, i.e. queries were expanded with a (varying) number of terms chosen by his term selection value (TSV). Then precision/recall statistics were obtained. Next, the HAL matrix for the relevant set was computed. For each query (1) the submatrix was taken containing only the terms of the expanded query, (2) this matrix was normalized to get a probability distribution, and (3) the stationary distribution was computed. This vector represented the expanded query. Finally, the cosine distance to the documents determined the new ranking. The procedure was repeated for query extension from 10 to 100 terms in steps of 10.
*Results.* The average precision for the stationary kernel

was a substantial improvement for all query extensions. The overall averages are summarized in Table 1.

## Conclusion

We started with the observation that in today's hard approach to information retrieval, more and more effort is needed to achieve even small improvements. Or, in economic terms, the approach seems to have reached a point of diminishing returns. As a complement to this approach we advocate a cognitive science approach, which is in the phase of increasing returns.

We described three cases in different areas of cognitive science, and showed that the principles that have been around in the field for a long time could be applied to compete with the prevalent hard approach of IR.

As may have transpired from the case descriptions, even the softer approach requires a substantial programming effort. The need to combine a thorough knowledge of psychology with good programming skills should make this approach particularly suited, attractive, and opportune for cognitive scientists. This kind of research is much needed in IR. And as information retrieval is an important pillar of the information society, we intend this paper also as a call to arms.

If that call is answered by cognitive scientists, we expect the results for information retrieval to be substantial, immediate, and fundamental.

## References


Adams, V. (1973). *An introduction to Moderm English word formation*: London: Longman.

Biederman, J. (1987). Recognition-by-components. a theory of image understanding. *Psychological Review*, *94*, 115–147.

Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, *25*, 211 – 257.

Croft, W. (1995). Effective text retrieval based on combining evidence from the corpus and users. *IEEE Expert*, *10*(6), 59–63.

Dumais, S. (2003). Data-driven approaches to information access. *Cognitive Science*, *27*, 491–524.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*: Cambridge: MIT Press.

Gardiner, D., Riedl, J., & Slagle, J. (1994). TREC-3: Experience with conceptual relations in information retrieval. In *Proceedings of the Third Text REtrieval Conference (TREC-3)* (pp. 333–352). Gaithersburg: NIST.

Hoenkamp, E. (1998). Spotting ontological lacunae through spectrum analysis of retrieved documents. In *Proceedings of the 15th European Conference on Artificial Intelligence ECAI-98. Workshop on Applications of Ontologies and PSMs* (pp. 73–77). Chicester: Wiley.

Hoenkamp, E., & de Groot, R. (2000). Finding relevant passages using noun-noun components. In M. Hearst, F. Gey, & R. Tong (Eds.), *Proceedings of the 23rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 385–387). New York: ACM.

Hoenkamp, E., Schomaker, L., van Bommel, P., Koster, C., & van der Weide, T. (1996). *Profile - A Proactive Information Filter* (Tech. Rep. No. 9602). Computer Science Institute, University of Nijmegen, the Netherlands.

Hoenkamp, E., Stegeman, O., & Schomaker, L. (1999). Supporting content retrieval from WWW via 'basic level categories'. In M. Hearst, F. Gey, & R. Tong (Eds.), *SIGIR '99: 22nd annual international ACM SIGIR Conference* (pp. 311–312). ACM Press, New York.

Ingwersen, P. (1999). Cognitive information retrieval. *Annual Review of Information Science and Technology*, *34*, 3–52.

Klein, W., & Perdue, C. (1997). The basic variety or: Couldn't natural languages be much simpler? *Second Language Research*, *13*(3), 301–348.

Lenat, D. (1995). Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, *38*(11), 33–38.

Levi, J. N. (1978). *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

Liddy, E., & Myaeng, S. (1993). Dr-link: A system update for TREC-2. In *Proceedings of the Second Text REtrieval Conference (TREC-2)* (pp. 85–100). Gaithersburg: NIST.

Miller, G. (1995). Wordnet: A lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Murphy, G., & Smith, E. (1982). Basic-level superiority in picture categorization. *Journal of verbal learning and verbal behavior*, *21*, 1–20.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The Measurement of Meaning*: Urbana: The University of Illinois Press.

Peterson, M., & Graham, S. (1974). Visual detection and visual imagery. *Journal of Experimental Psychology*, *103*, 509–514.

Robertson, S. E., Walker, S., Hancock-Beaulieu, M., Gull, A., & Lau, M. (1992). Okapi at TREC. In *Text REtrieval Conference* (pp. 21–30.

Rosch, E., Mervis, C. B., Gray, W. E., Johnson, E. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Ryder, M. E. (1994). *Ordered chaos: A Cognitive Model for the Interpretation of English Noun-Noun Compounds*. Ph.D. thesis, University of California, San Diego.

Salton, G. (1988). *Automatic text processing: the transformation, analysis and retrieval of information by computer*. Reading, Mass.: Addison-Wesley.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–9–5). Oxford University Press.

Tversky, B., & Hemenway, K. (1984). Objects, parts and categories. *Journal of Experimental Psychology*, *113*, 169–193.