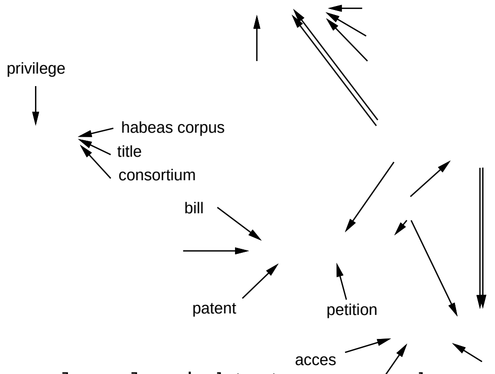

copyright
   the exclusive legal right to reproduce, publish,
   and sell the matter and form of a literary, musical,
   or artistic work -- copyright adj

license or licence
   1a permission to act
   1b freedom of action
   2a a permission granted by competent authority
      to engage in a business, occupation,
      or activity otherwise unlawful
   2b a document, plate, or tag, evidencing
      a license granted
   3a freedom that allows or is used with irresponsibility

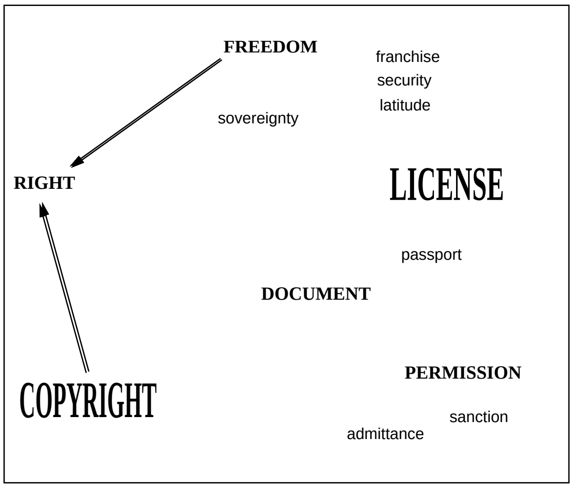

Figure 1: Simplified Semantic Net for "copyright" and "license"

Notice that "license" is a type of " permission" (W7:sense1a,sense 2a) and a type of "f
(W7: sense 1b). A user will be able to see this relationship in the browser, and pick othe
related to "permission" and "freedom" for searching, such as "franchise" or "access". A "
also a "document" (W7:sense 2b), and thus is related via the "document" concept to other wo
as "passport", "bill", "patent", "petition", "registration", and so on. In automatic se
higher in the net, those with more general meaning such as "action" or "thing" will be weight
since they tend to carry less specific content. This is the concept of a generality measu
from the net itself, based on frequency and net location. The impact of using such an appro
a key word search is that related concepts can be scanned for possible inclusion or exclus
query.

As demonstrated in Klavans et al. 1992, other complex relationships can be extracte
dictionaries, such as MEMBER_OF, TYPE_OF, GROUP_OF, QUALITY OF, CONSISTS OF, and so on. As
in the Figure, "copyright" and "license" are closely connected, via the "freedom" and "ri
Thus, links can be drawn between highly abstract concepts, such as "freedom", as well as
items, such as "document". Within the structure of the net, the hyponyms of "document" i
"patent", "bill", "certificate", "petition", and or course "license."

Our method combines statistical and symbolic approaches to the problem of text analys
purely statistical approach functions by converting words into numerical objects and proc
as any statistical data. This method is useful for handling large volumes of text, and has b
successfully in most IR systems. However, it fails to capture the fact that the lexicon ha
peculiar to language, and particular to the nature of linguistic objects. The purely symbo

[1] Portions of this paper arose from the Digital Library effort at Columbia University. In particular, I thank Kathleen McKeown, research director, for fruitful discussion of some of the ideas presented here. All inconsistencies or potential errors are, however, my own.

References

[1] Ahlswede, T., Anderson, J., Evens, M., Neises, J., Pin-Ngern, S., and Markowitz, J. “Automatic construction of a phrasal thesaurus for an information retrieval system from a machine readable dictionary”, Proceedings of RIAO, (1988), 597-608.

[2] Amsler, R., “Machine-Readable Dictionaries”, Annual Review of Information Science and Technology, (1984) 19, 161-209.

[3] Belkin N., “The Effect of Multiple Query Representations on Information Regtrieval System Performance”, in R. Korfhage, E. Rasmussen and P. Willer (eds.), Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (1993), 338-346.

[4] Chodorow, M., R. Byrd and G. Heidorn, “Extracting semantic Hierarchies from a large on-line dictionary”, in Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics, (1985), 299-304.

[5] Croft, W.B., H.R. Turtle and D. D. Lewis. “The Use of Phrases and Structured Queries in Information Retrieval”, in Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1991, 32-45.

[6] Evans, D. E., K. Ginther-Webster, M. Hart, et al. “Automatic Indexing using selective NLP and first-order thesauri” Proceedings of RIAO: Intelligent Text and Image Handling, 1991, 624-643.

[7] Fagan, Joel L. Experiments in Automatic Phrase Indexing for Document Retrieval: A comparison of Syntactic and Non-Syntactic Methods. PhD thesis. Department of Computer Science. Cornell University. 1987

[8] Fox, E. A. Tutorial on Knowledge-Based Information Retrieval. Fifteenth International Conference on Research and Development in Information Retrieval. Sponsored by the ACM SIGIR, 1992.