Commentary

In Due Diligence Searches For Prior Art,The Patent Office Simply Does Not Compute

***The Scientist*, Vol:9, #1, pg.13, January 9, 1995.**

Author: Eugene Garfield

Yogi Berra's classic line, "It's deja vu all over again," came to mind as I read a story on the United States Patent Office fiasco regarding the Compton's New Media case (S. Chartrand, "At the Patent Office, a digital dawn," New York Times, Nov. 12, 1994, page 39).

To recap: In August 1993 the Patent Office awarded Compton's, a San Diego-based CD-ROM publisher, exclusive rights on a basic search and retrieval software feature in its interactive CD-ROM products. This would have allowed Compton's to demand licensing or royalty fees from virtually any company in the CD-ROM industry. However, last October the Compton's patent was invalidated. After a more careful examination, the Patent Office found dozens of documents indicating the "invention" was neither new nor unique. It is indeed ironic that this obvious failure in searching "prior art" involved a retrieval software patent application.

The sense of deja vu arises from my own experience with the Patent Office and its efforts at computerization more than 35 years ago. In 1957, the office was swamped by a flood of patent applications on steroids, then a very active field of research and commercialization. In order to alleviate the huge backlog, the Pharmaceutical Manufacturers Association (PMA) negotiated and financed a steroid literature coding project with the Patent Office. A simple system of encoding information for each new steroid compound was devised and an outside contractor was selected to screen the current literature for new compounds. IBM punched cards--the state of the art at the time--were used to store the coded information.

The outside contractor was Eugene Garfield Associates, an information-engineering consulting firm I founded in 1954 and the predecessor of the Institute for Scientific Information (ISI). Over the short three-year life of this project, more than 11,000 steroid compounds were encoded. The project experience led to the establishment of ISI's Index Chemicus Registry System. Launched in 1960, Index Chemicus has gone on to cover more than 6 million new compounds.

Before the advent of Index Chemicus, few people believed my assertion that an up-to-date molecular formula index to the literature could be created. It had always been assumed that one first had to name compounds before they could be indexed. Chemical nomenclature was--and still is--an arcane linguistic skill.

Parenthetically, I described the generative grammar of Geneva chemical nomenclature, which some may now call the IUPAC system, in an appendix to my doctoral dissertation on computer algorithms for translating chemical names into molecular formulas, which was summarized in Nature (E. Garfield, Nature, 192:192, 1961).

The Patent Office should keep in mind the importance of linguistic nuances as it pursues its long-overdue campaign to computerize its operations. Like so many other institutions, the Patent Office is facing a massive textual search problem. Investment in computer hardware and software alone will not solve the problem, the Patent Office must come to grips with the ability of inventive people to express similar ideas in an endless variety of terms. It is not just a matter of searching full texts, but rather the attempt to deal

with meaning -- to translate one linguistic expression into another.

When patent examiners are researching a new patent application, a simple vocabulary-based search strategy would be inadequate to identify prior art described in different terms but related to the same basic concept. Even with the help of the Patent Office's many subject specialists, key relevant prior art may continue to be overlooked.

In recognition of the limitations of title, keyword, and human indexing, ISI developed the Related Records hypersearch option and KeyWords Plus. These innovations, respectively, link related documents through the references they cite and enhance their description by adding keywords derived from the titles of cited references. As has been frequently demonstrated, documents identified by bibliographic coupling (Related Records) or KeyWords Plus prove to be highly relevant although they may not be retrieved by title word searches.

The Patent Office would be well advised to look beyond computer hardware and software as a solution to its backlog and prior art searching problems. It should plan from the start to also develop the ability to link patents by conceptual meaning, not simply textual vocabularies. Many of these conceptual links can be found via the references cited in each patent application. But if the Patent Office relies solely on a digital technology crutch, it may well be "deja vu all over again.". .. . . . .