Using KOS as a basis for text analytics and trend forecasting.

This case study in term analytics takes the approach of text analytics but raises the standard by first extracting conceptual information using a KOS or thesaurus. This approach is used on ten years of publishing proceedings, journal articles conference papers and other materials held in the Institute for Electrical and Electronics Engineering (IEEE) Xplore data base. From this data we are able to establish the trends for their publishing over ten years. We took ten years of Xplore data and indexed it with the revised IEEE thesaurus. We overlaid the source society and 16 major categories to group the data. We then indexed the same ten years of data using the US National Library of Medicine MEsH and US Defense Technical Information Center (DTIC) thesauri to see where the overlap edges of the science are. Finally we took the United States Patent and Trademark Office data for the same ten years and indexed it with the same thesauri to see how patent and innovation trends lead or followed the publishing by proceedings and by journal articles. The results show the IEEE where their societies are heading and what the trends are at a very discrete level.

Marjorie M.K. Hlava
Access Innovations / Data Harmony
www.dataharmony.com
www.accessinn.com