

8-16-1996

Is Relevance Relevant: The History and Effect of the Relevance Measure on IR

Judith Pinn Carlisle

Georgia Institute of Technology, judith.carlisle@mgt.gatech.edu

Follow this and additional works at: <http://aisel.aisnet.org/amcis1996>

Recommended Citation

Carlisle, Judith Pinn, "Is Relevance Relevant: The History and Effect of the Relevance Measure on IR" (1996). *AMCIS 1996 Proceedings*. 189.

<http://aisel.aisnet.org/amcis1996/189>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISEL). It has been accepted for inclusion in AMCIS 1996 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

Is Relevance Relevant: The History and Effect of the Relevance Measure on IR

by [Judith Pinn Carlisle](#)
Georgia Institute of Technology
Atlanta, GA 30332-0520
judith.carlisle@mgt.gatech.edu

Automated information retrieval system (IRS) experimentation and evaluation emphasizes two measures --- recall and relevance (or precision). [van Rijsbergen, p. 32] defines these measures thus:

"...precision being the proportion of the retrieved documents that are relevant, and recall being the proportion of the relevant documents that have been retrieved." [, p.32]

which immediately leads to two questions, first, what does 'relevance measure' measure, and second, how is the relevance measure derived? This paper explores the history and evolution of the relevance measure and assesses its impact on IRS evaluation.

Relevance measures how well a retrieval set fulfills the information needs and expectations of the querier. However, until such needs can be effectively represented in a computational environment, they remain elusive. Inclusion of measurement of such needs in the realm of information science requires the inclusion of such areas as psychology, cognitive science, and artificial intelligence. Further, the focus of information research moves from ascertaining how retrieval sets are formed toward ascertaining how individuals respond to retrieval sets. Therefore, inclusion of the relevance measure in IRS evaluation requires that topics not necessarily pertinent to the retrieval function be included in the research field.

The relevance measure has the following characteristics. It is:

- 1) a subjective measure,
- 2) derived from behavioral observations of experimental participants, and
- 3) unable to assess automated text retrieval as a deterministic function.

Hence, the relevance measure is not a useful measure when ascertaining IRS performance and should be reconsidered or reconstructed.

On Measurement

Lord Kelvin is quoted as saying: "I often say that when you can measure what you are speaking about, you know something about it; when you cannot measure it ...your knowledge is of a meagre and unsatisfactory kind." [Thompson, p. 73] Information sciences more and more, attempt to quantify natural phenomena and events in order to describe them more

precisely. While all measurement is dependent upon the measurement tool, two major approaches to measurement can be identified, absolute and relative. Absolute measurement relies on a predefined standard and evaluates items in accordance with this standard. Relative measurement evaluates an item by

determining its relationship to another item. Relative measurement is useful only when the reference tool is known and understood to others outside of the experimental environment.

Relative and subjective approaches to measurement must be carefully assessed as the measurement process is not independent from the measurement tool. Both the relevance and recall measure are subjective measures. Research such as [Burgin] shows little variation between user-based assessment of relevance, such research fails to take into account various types of IRS user. Still, research currently underway finds significant evidence that executive users have very different desires with regard to retrieval set characteristics than, say, lawyers. Such differences with regard to user type and retrieval set assessment lends credence to pre-designation of IRS user classes, and calls into serious question the usefulness of the relevance measure.

On Relevance

The relevance measure is found in the earliest of IRS texts as [Swanson], and [Salton]. However, This measure evolved throughout the 1970s and by the 1980s had subsumed the "pertinence" measure becoming one of the two dominant evaluative measures for IR. Current IRS evaluation continues to use the relevance measure.

Three measures dominated early IRS evaluation, pertinence, relevance, and recall. With regard to a retrieval set obtained by an IRS, in response to a user-supplied query, these three measures can be described thus. Pertinence measures how many documents in a retrieval set are useful to the querier. Relevance determines whether or not each member of the retrieval set contains the information concept intended to be encoded into the query by the user. Recall measures how many documents in a document base which contain an information concept are obtained by the retrieval function. Starting with the thesis that "different views of relevance arose because relevance was considered at a number of different point in the process of knowledge communication," [Saracevic] presents a thorough, yet concise, evaluation on changing notions of the relevance measure and allows us to ascertain that the measure now known as relevance has in fact, subsumed the pertinence measure. We believe that the relevance measure has evolved to its current state as a result of the lack of empirical differentiation between user intent and system performance.

Expressive Capacity and Information Retrieval

Expressive capacity of a language "L" shall be described as L's ability to express (and recognize) an infinite number of meaningful strings. In this paper, we correlate the notion of a meaningful string with that of a user-supplied query. Though much expressive capacity is required to encode, and differentiate between, nuances of language, increases in expressive capacity also allow for more ambiguity in language, allowing the same meaningful expression to have more than one meaning. While some philosophies of language, such as [Frege]'s, view such events as imperfections in human language others [Donnellan], [Kripke], [Putnam] assess how human use of language allows humans to often disambiguate such expressions. Implicit in each of these philosophies is acknowledgement of intent when a concept when a

concept is encoded into a meaningful expression.

Unfortunately, machines cannot recognize intent. Thus, when a machine interprets a user-supplied query (that is obtains a retrieval set) the machine interacts only with its own understanding of that query. The retrieval function alone determines the range of the retrieval set as expressed by the query. User-assessment of retrieval sets (i.e. the relevance measure) fail to view automated retrieval in such a straight-forward manner, and result in poor measures of relevance.

We describe the retrieval function via three cooperative aspects:

- 1) the user-supplied query,

- 2) the surrogate document base representation, and
- 3) the function which matches similar documents to the query.

As such, the function is deterministic and independent of any desires or beliefs a user may intend when composing a query. The inability of a particular IRS to successfully include user-intent into the retrieval function results from lack of expressive capacity of the query language, and also the surrogate document representation. Further, typical of all computer-based ISes, the system operates in a precalculated, deterministic manner. Therefore, two users, each with distinct and unique information needs, may create and submit to the IRS the same query.

Many researchers [Smeaton], have recently incorporated syntactic processing techniques into the development of a surrogate document representation. Still, due to their focus on syntactic, rather than conceptual, representation, syntactic approaches alone cannot be expected to significantly increase IRS ability to account for user expectations.

Conclusion

Useful assessment of IRS performance is negatively impacted by the relevance measure. Much of the nature of the retrieval function itself is not yet well understood. Even less has been ascertained about the manner in which different retrieval functions obtain retrieval sets. Both of these topics are

important to explore more thoroughly. Only when the function itself is understood, will successful manipulation of retrieval set creation be achieved.

Though it is essential that users be satisfied with system performance, continued use of the relevance measure shifts the direction of IRS research from understanding the retrieval function to attempting to fulfill the unmeasurable desires of users. Further, The relevance measure causes users to expect unreasonable performance from IRSes.

Bibliography

Burgin, Robert. "Variations in relevance judgments and the evaluation of retrieval performance," *Information Processing and Management*, vol. 28, no.5, 1992, pp. 619-627.

Donnellan, K. "Reference and Definite Descriptions," originally published in 1966, reprinted in Harnish, Robert M., (ed) *Basic topics in the philosophy of language*, Prentice Hall, New Jersey, 1994.

Frege, G. "On Sense and Reference," in Geach, P. and Black, M. *Translations From the Philosophical Writings of Gottlob Frege*, Second Edition Blackwell Publishing, 1970, pp. 56-78.

Kripke, Saul A. *Naming and Necessity*, Harvard University Press,

Putnam, Hillary. "The meaning of 'meaning,'" originally published in 1975, reprinted in Harnish, Robert M., (ed) *Basic topics in the philosophy of language*, Prentice Hall, New Jersey, 1994.

Saracevic, Tefko. "Relevance: A review of and a framework for the thinking on the notion in information science," in *Journal of the American Society for Information Science* --- November-December, 1975, pp. 321 - 343.

Salton, G. *Automatic Information Organization and Retrieval*, New York: McGraw-Hill, 1968.

Smeaton, Alan F. "Natural language processing and information retrieval," *Information Processing and Management*, vol. 26, no. 1, 1990, pp. 19-20.

Swanson, Don R. "Searching natural language text by computer," in *Science*, vol. 131, October 21, 1960, pp. 1099 - 1104.

Thompson, William (Lord Kelvin). "Electrical units of measurement," in *Popular Lectures and Addresses*, London, 1989, vol. 1.

van Rijsbergen, Cornelis J. "Retrieval Effectiveness," in Sparck-Jones, Karen (ed). *Information Retrieval Experiment*, Butterworths, London, 1981.