

"In the catalogue ye go for men": evaluation criteria for information retrieval systems.

[Julian Warner](#)

School of Management, The Queen's University of Belfast, Belfast,
BT7 1NN, Northern Ireland

Abstract

The contrast between the value placed on discriminatory power in discussions of indexing and classification and on the transformation of a query into a set of relevant records dominant in information retrieval research has not been fully explored. The value of delivering relevant records in response to a query has been assumed by information retrieval research paradigms otherwise differentiated (the cognitive and the physical). Subsidiary concepts and measures (relevance and precision and recall) have been increasingly subjected to critiques. The founding assumption of the value of delivering relevant records now needs to be questioned. An enhanced capacity for informed choice is advocated as an alternative principle for system evaluation and design. This broadly corresponds to: the exploratory capability discussed in recent information retrieval research; the value of discriminatory power in classification and indexing; Giambattista Vico's critique of the unproductivity of Aristotelian methods of categorisation as routes to new knowledge; and, most significantly, to ordinary discourse conceptions of the value of information retrieval systems. The criterion of enhanced choice has a liberating effect, restoring man as an artificer and enabling a continuing dialectic between theory and practice. Techniques developed in classic information retrieval research can be adapted to the new purpose. Finally, the substitution of the principle of enhanced choice exemplifies the development of a true science, in which previous paradigms are absorbed into new as special cases.

Ay, in the catalogue ye go for men;
As hounds, and greyhounds, mongrels, spaniels,
curs,
Shoughs, water-rugs, and demi-wolves, are clept
All by the name of dogs: the valu'd file
Distinguishes the swift, the slow, the subtle,
The housekeeper, the hunter, every one
According to the gift which bounteous Nature
Hath in him clos'd; whereby he does receive
Particular addition, from the bill
That writes them all alike;

Shakespeare. *Macbeth*. c.1606.

Introduction

The epigraph indicates the value which has been historically attached to subtlety of distinctions in the language or lexicon of information retrieval systems. In this respect, the passage anticipates the principle formulated in modern discussions of indexing and classification that the value of an index term lies in its discriminatory power. In this principle, and in its historical anticipation, there is a strong, although largely unnoticed, contrast with the assumption of information retrieval research, particularly experimental information retrieval research, that the

performance of an information retrieval system is to be measured by its capacity to deliver relevant records in response to deliberately articulated queries.

The concern here is not, then, with the uses of classification in information retrieval but with the broader question of whether the central principle embodied in the practice and theory of classification and indexing can yield more satisfying design and evaluative criteria for information retrieval systems than those which have been characteristically assumed in information retrieval research. Two paradigms have been distinguished in information retrieval research, the cognitive and the physical, but they share the assumption of the value of delivering relevant records ([Ellis 1996](#): 19; [Belkin and Vickery 1985](#): 114). For the purposes of the discussion here, they can be considered as a single, if heterogeneous, paradigm, linked if not united, by this common assumption.

The contrasting paradigm implicitly embodied in classification and indexing may finally be incommensurable with that of information retrieval research, with disputes not logically resolvable within either paradigm. The approach taken in this paper will be suggest:

- that an alternative principle involving discriminatory power has been held, implicitly and explicitly, in a number of largely separate discourses;
- that the cumulative effect of recognising this is to indicate more viable and productive criteria for designing, using and evaluating information retrieval systems;
- and, finally, that the classical tradition of information retrieval research can itself be assimilated to the new model.

In this final respect, the development proposed here is an exemplar of scientific development in which discarded paradigms are absorbed into developing ones, as special cases.

The discourses in which an alternative principle for the design and evaluation of information retrieval systems can be discovered and which are to be covered here are:

- an emerging, although rather isolated and discontinuous, strand of information retrieval research;
- accepted discussions of the principles of classification and indexing;
- Giambattista Vico's critique of Aristotelian principles and categories for classification;
- and, most crucially, ordinary language discussion of information systems.

The paper will review information retrieval research, taking the liberty of conflating distinguishable aspects for the purposes of a higher level discussion, and then indicate that an alternative principle for evaluation can be found in the discourses identified. The value of the alternative model developed will be explored. In conclusion, it will be suggested that the alternative principle and criteria developed can have liberating effect, allowing theory and practice to interact, and that a productive transformation of the field has been indicated.

Alternative paradigms

Information retrieval research

Information retrieval research, particularly in the experimental tradition emerging in the 1950s in Britain and North America, has taken as its founding assumption the principle that an ideal information system should deliver all (and possibly only all) the records relevant to a given information need. In order to evaluate information systems in relation to this desired end, or variations on it, various steps were taken: relevance was stabilised and quantified, sometimes being reduced to a binary or dichotomous variable; and measures of precision and recall, which depend on the prior stabilisation of relevance, were developed. More recent research has questioned the validity of aspects of this paradigm, although more frequently with reference to its subsidiary concepts (relevance and information need) and derived measures (precision and recall) than with regard to its founding assumption.

The adequacy of the concept of relevance employed information retrieval research has been questioned. Experiments substitute a measurable phenomenon, relevance as constructed under artificial conditions, for an unmeasurable one, relevance under operational conditions, but fail to demonstrate that there is an adequate correlation between the two. Most disturbingly, it has been suggested that operational relevance is fluid, influenced by intention, context, and other documents seen or read, and simply not amenable to stabilisation or, further, quantification ([Ellis 1984; 1996](#)).

The classical measures of precision and recall are also rendered increasingly artificial by the high degree of interactivity enabled by recent information technology developments. How, when searching a CD-ROM database, is the final set of records to be isolated except by a process whose very arbitrariness invalidates it as a component of a measure of system performance? ([Warner, 1992](#)) High interactivity, and unmediated searching, also reduce the need for a query to be fully articulated in advance of searching. There has also been a realisation that a deliberately stated query (which can be distinguished from an information need or assertion of relevance) may be a methodological requirement for controlled experiment, but is not intrinsic to the information seeking situation ([Heine, 1977](#)) and that it is possible to search without verbalising an information need.

The classic information retrieval paradigm, and the concepts and measures associated with it, could be preserved but only at the cost of increasing its distance from more realistic information seeking situations. It may be that not only are the classical concepts and measures both becoming and being recognised as increasingly artificial but that the founding assumption - that a system should deliver all (and only all) the relevant records should be re-examined. What is required, then, is not questioning of concepts with the paradigm but of its founding assumption, turning what has been received as a given into an object of enquiry.

To some extent, this has begun to occur within information retrieval research. The subtlety and complexity of information retrieval has been recognised (Swanson 1988). Most specifically, the principle of exploratory capability, the ability to explore and make discriminations between representations of objects, has been suggested as the fundamental design principle for information retrieval systems ([Ellis 1984; 1996](#)).

On a subjective level, this can be supported by introspection: that what I desire from an information retrieval systems is not a possibly mysterious transformation of a query into a set of records, but a means of enlarging my capacity for informed choice between the representation of objects within the given universe of discourse. Such an enhanced capacity for informed choice broadly corresponds to exploratory capability. It could also be regarded as analogous to a sense of cognitive control over, or ability to discriminate between, representations of objects.

One example (which may be fictional in a double sense) can be given of the need for enhanced discriminatory power. At one point in time, a researcher might wish to distinguish the private individual, Samuel Langhorne Clemens, from the author, Mark Twain, (perhaps out of interest in his copyright disputes or in his brother's, Orion Clemens, activities as Secretary to Nevada Territory). What would be valuable for this purpose would be a system which did not conflate these two distinguishable aspects of the individual but enabled them to be differentiated. At a later point in time, the same researcher might be interested in information on Mark Twain and Samuel Clemens considered as single entity. An information retrieval system should then be able to differentiate and to link together the occurrences of these different names, as required.

In conclusion, the assumption that it is desirable to obtain all, and possibly only all, the records relevant to a given query can be rejected in favour of the alternative principle of exploratory capability or enhanced capacity for informed choice. Introspection supported the value of exploratory capability. Its appeal as an alternative to the established information retrieval paradigm could be strengthened if it could be found, even if only implicitly or in analogous forms, in other, independently developed, discourses.

Principle of indexing and classification

An acknowledged principle of indexing and classification is that the value of a term is its discriminatory power. By discriminatory power is understood the ability to partition and select from the objects represented within the given universe of discourse. What particular terms or methods of classification are appropriate will vary with the area of discourse and the focus of interest: most obviously, a factor which differentiates one set of objects from another will not serve to discriminate within either set of objects. Discriminatory power is again analogous to exploratory capability, or, more accurately, a critical factor in enabling progressive and controlled exploration.

Vico's critique of Aristotelian classification

A strong, and highly significant, analogue to exploratory capability can be found in Vico's critique of Aristotle. Aristotle's philosophy, as well as being a direct and indirect source for subsequent understandings of genus, species, specific difference, synonymy and equivalence, involved, in some of its aspects, a systematic method of enquiry in order to classify an object. An enquirer was required to ask a series of questions: Does the thing exist? What is it?

How big is it? What is its quality? and the like. This method of enquiry was subjected to an incisive critique by Vico:

Aristotle's *Categories* and *Topics* are completely useless if one wants to find something new in them. One turns out to be a Lull or Kircher and becomes like a man who knows the alphabet, but cannot arrange the letters to read the great book of nature. But if these tools were considered the indices and ABC's of inquiries about our problem [of certain knowledge] so that we might have it fully surveyed, nothing would be more fertile for research. ([Vico 1988](#): 100-101)

The last clause of that critique deserves emphasis, 'nothing would be more fertile for research.' The rigidity of the method is avoided, while some of its techniques are retained, and it is transformed into a systematic and effective means for enhancing knowledge of an object. Analogously, while rejecting the rigid transformation of a query into a set of records assumed as desirable in information retrieval research, similar techniques can be used to explore the domain of discourse covered by the information retrieval system.

A further supporting analogue can be found in the fictional rather than discursive treatment of rigid classifications in Dickens' *Hard Times*. The logical distinctions exemplified in Bitzer's definition of a horse - 'Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth, and twelve incisive ... Age known by marks in mouth.' (Dickens 1989: 6) - which does resemble 19th century taxonomies for the horse, themselves influenced by the Aristotelian method of definition by genus and species, are presented as harsh ([Warner 1994](#): 106-108). Outside the restricting enclosure of the town, a different metaphor for knowledge is discernible:

They walked on across the fields and down the shady lanes, sometimes getting over a fragment of a fence so rotten that it dropped at a touch of the foot, sometimes passing near a wreck of bricks and beams overgrown with grass, marking the site of deserted works. They followed paths and tracks, however slight. ([Dickens 1989](#): 353)

The value of an information system could then be the ability it offers discriminatingly to follow 'paths and tracks, however slight'. Classification schemes themselves (and their analogues in thesaural relations among indexing terms) can then be received not as fixed models of stable entities but as valuable exploratory devices.

Ordinary discourse

Ordinary, particularly informal spoken, discussion of information systems is simultaneously highly significant and difficult to produce as evidence. Evaluative criteria may be implied rather than explicitly articulated. Yet when a searcher complains that it is difficult to control the number of records retrieved, a principle of discriminatory power is being invoked. More explicitly, one spoken response to an earlier version of this paper was: 'that's the basis [an enhanced capacity for informed choice] on which people use systems anyway'.

Summary

Similarities in themes and principles enunciated or implied have been revealed in largely separate discourses, emerging in information retrieval research, implied in discussions of principles of indexing and classification, made explicit in Vico's critique of Aristotelian methods of investigation, and present, in partly unarticulated form, in ordinary discourse. The mode of expression varies, but an enhanced capacity for informed choice, for effective discrimination, or for cognitive control was discovered to be valued in all the discourses adduced. Independent agreement with an emerging and rather isolated theme in information retrieval research, of exploratory capability, offers supports for replacing the established emphasis on the delivery of relevant records with such a principle for the design and use of information systems. In some respects, possibly through the influence of concepts of classification and of ordinary discourse understandings, working systems may offer exploratory capability and productive interaction. In Vico's terms, practical understanding has been in advance of theoretical articulation.

Evaluative model

Endorsing the principle of enhanced capacity for informed choice can have a liberating effect, revealing the intra-theoretic nature of many disputes within the classic tradition of information retrieval research: it offers the possibility of a deeper understanding of relevance; enables a mutually informing relation between practice and

theory; restores man as artificer as a designer and user of information systems rather the cipher of information retrieval research; and can enable the development of more satisfying evaluative criteria.

Disputes over the validity of constructs demanded for retrieval system evaluation in the classic tradition of information retrieval research, for instance whether deliberately contrived relevance judgements are adequately correlative with real world judgements, can now be regarded as intra-theoretic, connected with the theoretical framework imposed, not inherent in the process of information retrieval and not necessarily contributing to an understanding of those processes. In some respects, the construction imposed by the research paradigm may even have inhibited development of understanding of its chosen domain of study. For instance, the methodological need to reduced relevance to assessments, possibly open to quantification, and stable over time, may have inhibited exploration of its many possible dimensions. Some dissenting discussions have insisted on its complex and multi-faceted nature ([Wilson 1973](#); [Watson, et al., 1973](#)).

A mutually informing and productive relation between theory and practice can be developed. For instance, the practical experience of those indexing procedures or retrieval algorithms which enhance exploratory capability in specified circumstances can inform theoretical development and system design and modification. The divorce of information retrieval research from practice has been noted and sometimes regretted, although less often explained. Now the practical understanding embodied in working systems can be recognised and theoretically developed.

The further question then also arises as to whether accepting the principle of exploratory capability has practical implications in terms of the indexing procedures or algorithms for searching to be adopted. An immediate response would be that it does not necessarily have unambiguous practical implications: That the particular indexing procedures and algorithms to be used will be critically dependent on the purpose and context of retrieval, including the cost of indexing and retrieval. Crucially for continuity of systems development, techniques identical with or analogous to those currently developed may be used to different ends. It should also be noted that the Boolean logic used in many retrieval systems, does, under conditions, have the advantage of relative transparency to the searcher. The objection that it is an ineffective way of transforming an information need into a set of relevant records is no longer tenable. It could still be objected that in some applications, for instance with heterogeneous textual material without humanly assigned index terms, it gives inadequate control over the representations within the universe of discourse.

A deeper effect is to restore man as an artificer and to recognise the subtlety of the processes involved information retrieval. Rather than being subjected to retrieval process beyond immediate control, the searcher is presented with an enhanced capacity for choice and for making recalled sets. The new, and historically unprecedented, potential for enhanced forms of knowing of existing textual material can then productively explored. For instance, the unrivalled opportunity offered by full text database for exploring the semantic mutability of written word forms with different contexts can be pursued.

More detailed evaluative criteria could be developed from the central evaluative principle of enhanced choice, partly by drawing on the understandings already developed in discussions of classification. Yet it should be recognised that quantitative comparative measures are unlikely to result. Once the diversity of contexts for information retrieval is recognised, the idea of a single, generally applicable approach to system design, or a single comparative measure of system performance, becomes severely questionable. The best outcome which can reasonably be expected from research and from reflection on practice is a better understanding of the process of information retrieval, which can then be used either to design better information systems or to make more effective use of existing systems.

Conclusion

Replacing the emphasis on the delivery of relevant records with a stress on exploratory capability or cognitive control as a design and use principle for information retrieval can have a liberating effect. It yields more satisfying evaluative criteria while preserving a strong continuity with previous work, particularly in recognising the utility of developed information retrieval techniques. Theory and practice, rather than being separate or even antagonistic, are enabled to inform each other. The discourses, of philosophy, classification and ordinary discussion, from which the new principle has been drawn, can be brought further to bear upon information retrieval, transforming it into a human science and recognising its subtlety and significance. A minor, although significant, relief, is liberation from the obligation to read work in the classic information retrieval paradigm, except for the emerging signs of self-questioning.

The transformation advocated in this paper resembles, in some respects, a mathematical revolution and can also be seen as an example of scientific progress. Classically fundamental transformations of mathematics have preserved the form while modifying the interpretation ([Ramsey 1990](#)); analogously, information retrieval techniques have been preserved but adapted to a new end. More broadly, it has been argued that a discipline exhibits the history of a true science if its earlier stages can be seen as special cases, from the perspective of its subsequent development ([Roberts 1982](#)): in this context, the automatic transformation of a query into a set of records can be seen as a possible support for informed choice, valuable in certain sets of circumstances. Rather than, as [Swanson](#) (1988) indicated, 'Waiting for Godot [while failing] to grasp what is now within reach', we can begin to explore the potential for improving human interaction with recorded knowledge.

References

- Belkin, N.J. and Vickery, A. (1985). *Interaction in information systems: a review of research from document retrieval to knowledge-based systems* (Library and Information Research Report 35). London: British Library, 1985.
- Dickens, C. (1989). *Hard Times*. Edited with an Introduction and Notes by Paul Schlicke. Oxford and New York: Oxford University Press, 1989.
- Ellis, D. (1984). Theory and explanation in information retrieval research. *Journal of Information Science*. 8, 1984, pp.25-38.
- Ellis, D. (1996). *Progress and problems in information retrieval*. London: Library Association, 1996.
- Heine, M.H. (1977). The 'question' as a fundamental variable in information science. In: O. Harbo and L. Kajberg editors. *Theory and application of information research* (Proceedings of the Second International Research Forum on Information Science, 3-6 August 1977, Royal School of Librarianship, Copenhagen). London: Mansell, 1980. pp.137-145.
- Ramsey, F.P. (1990). The foundation of mathematics. In: D. H. Mellor editor. *Philosophical papers*. Cambridge etc.: Cambridge University Press, 1990. pp.164-224.
- Roberts, N. (1982). A search for information man. *Social science information studies*. 2, 1982, pp.93-104.
- Swanson, D.R. (1988). Historical note: information retrieval and the future of an illusion. *Journal of the American Society for Information Science*. 39, 2, 1988, pp.92-98.
- Vico, G. (1988). *On the most ancient wisdom of the Italians: unearthed from the origins of the Latin language: including the disputation with the Giornale de' letterati d'Italia*. Ithaca and London: Cornell U.P., 1988.
- Warner, J. (1992). Retrieval performance tests in relation to online bibliographic searching. In: Deborah Shaw editor. *ASIS '92: Celebrating change: information management on the move* (Proceedings of the 55th ASIS Annual Meeting, Pittsburgh, PA, 26-29 October 1992). Medford, NJ: Learned Information for the American Society for Information Science, 1992. pp.231-241.
- Warner, J. (1994). *From writing to computers*. London and New York: Routledge, 1994.
- Watson, L.E. and others. (1973). Sociology and information science. *Journal of librarianship*, 5, 1973, pp.270-283.
- Wilson, P. (1973). Situational relevance. *Information storage and retrieval*. 9, 8, 1973, pp.457-471.

Acknowledgement

Another version of this paper is to be published as: "Can classification yield an evaluative principle for information retrieval systems?" In: Rita Marcella and Arthur Maltby editors. *The future of classification*. London: Gower, (in press).

How to cite this paper:

Warner, Julian (1999) ""In the catalogue ye go for men": evaluation criteria for information retrieval systems." *Information Research*, 4(4), paper 62. Available at: <http://informationr.net/ir/4-4/paper62.html>

© the author, 1999. Last updated: 24th June 1999

Check for citations, [using Google Scholar](#)
