Information Retrieval System Design for Very High Effectiveness

Joel Stubinz, PhD Samuel Whighli, FNH
Division of Computer Science, Endeavour Research and Development (BVI)
Suite 180, 85 Grattan St., Carlton 3053 Australia
joelstubinz@hotmail.com, s9800123B@bigfoot.com

Summary

Much of the research in Information Retrieval has concerned improvements to similarity computations, statistics gathering, and term extraction, with the goal of improving effectiveness. However, a simple examination of user characteristics can readily show, the method of computing similarity is less important than the behaviour of the system interface and environmental factors. It is hypothesised that there must be knowledge of the relationship between a query, its user, the environment, and the query and user instantiation in the real world. This hypothesis and others are demonstrated. With facilities for interaction and feedback appropriately incorporated, effectiveness of 100% can be achieved.

1 Introduction

Information Retrieval is the science of locating, from a large document collection, those documents that fulfil a specified information need [1, 2, 3, 4]. Much of Information Retrieval research is concerned with proposing and testing of methodologies intended to perform this function. To perform such tests it is necessary to make assumptions about the behaviour of users and the properties of text. For reasons of experimental design (following the assumption that "good" experiments should not have lots of variables) the user is often assigned the role of reader with no part in the process that produces answers from the document collection.

It might be thought that a formal model of the relationships between queries, documents, meaning, and relevance could be used as a foundation for information retrieval. It is argued that there can be no such model, humans cannot be left out of the equation yet cannot be modelled. (This paper does not consider the information needs of non-humans, such as robo-cup competitors.) This paper considers the basis and aims of information retrieval, examining assumptions and, on the basis of these observations, describes user experiments showing just how much effectiveness can be improved. These experiments justify great optimism for future system measurement and design, with full or at least 100% effectiveness easily achieved.

Language and text and their impact on information retrieval are considered first, then there is examination of the interaction of users, their environment, and relevance. The suggested system design and experiments are then reported. After everything is presented, a conclusion follows.

2 Similarity and information retrieval

It is well known that languages evolve over time [5, 6]. Our language is not the same language written by Shakespeare, nor was his the language of Chaucer. Even over periods as short as years there are trends and changes in English as words acquire new meanings and senses, old forms become obsolete, and the language of technology becomes part of the everyday patios [7].

It is not just words that change, but also the ways in which they are aggregated into units of larger meaning. For example, scientific papers from the turn of the century are by today's standards turgid and impenetrable [8, 9]; the development of modern journalistic style, itself a byproduct of the constraints of the telegraph, has had an enormous impact on scientific writing. Style has become simpler, more accessible, even light and humorous [10]; and at the same time use of language has become more sophisticated, in keeping with the growing erudition of the population at large. Even more surprising is the degree of evolution in popular writing, journalistic and otherwise, over the last thirty years. Newspaper articles from the 1960s use idioms and forms that already seem archaic and stilted. What, for example, would be made today of a report of woman being tried for murder, described by the defence counsel as "not of an unnatural nature" [11]? Yet this was the phrase of choice not long ago. The development of electronic word-processing, rapid dissemination of text, and reduced editorial control over the printed word, all contribute to rapid ongoing change in written English. The global village of the Internet, for example, has seen the birth of new word-bites to express happiness, sadness, and issues of reading details in manuals [12].

Of interest is in the impact of language evolution on information retrieval. It is likely that the changes in language operate at many levels—not just of word use and sentence construction, but at the level of epistemology and the interpretation of knowledge [13]. Change in the understanding of the meaning of knowledge would imply that the basis of information retrieval itself was unsound. Results could not be compared across time, nor would old databases be valid as a current research tool.

Remarkably, the existence of flux in understanding of knowledge can be explicitly tested. The authors' research has led to the observation that similarity is cultural. At the beginning of automated information retrieval, some forty years ago, Boolean queries were an accurate reflection of user's information needs [14]. But language and needs outmoded this approach, leading to the need for approximated similarity measures such as the inner product, the cosine measure, and probabilistic retrieval [15]. For many years these were the methods of choice, but were based on studies that neglected the user, and now the possibilities have been completely explored. Having under gone heavy experimental work investigating likely limits in all methods seen, this was not unlike a cryptogram. It has also been observed that the system in which the measure is implemented is also important, thus indicating that the methods of information retrieval are dynamic at all levels. In addition, semi-colons and other punctuation are in prolific use in recent scientific writing; it is likely that such trends will continue.

Another recent development is the successful application of natural language processing. For English as it was used ten years ago, experiments using the semantics of text were a failure. But important new experiments reverse this trend [16]. This is because improvements in education mean that more people can distinguish syntax from semantics. Moreover, successes with

techniques based on phonetic similarity are just in, so believers claim.

Information retrieval techniques must, it follows, account for language, culture and behaviour. Once for example estimation of similarity was always circumscribed and bounded, as in the cosine measure. But for browsing and learning this limiting of the user is not appropriate. A new method is the tangent measure [17], which has the peculiar advantage that relevance is unbounded. The rise of the tangent measure is due to two factors—the tendency of modern writing to be digressive, and the recent belief that answers should pose further questions. Also, the further the user is led from the topic of the query, the more knowledge is acquired, surely a positive feature of any information system.

In summary, it seems that the simple relationship between document and query, so widely assumed, must be at least questionable. A procedure for identifying purportedly relevant documents will only be valid in the context of one culture, one era. Even the notion of relevance is slippery, as is now discussed.

3 Users and relevance

Relevance judgements must embody the phenomenology of the user. It has widely been assumed that the user has a precise informational need, of which the query is an informal expression [18]. The information retrieval mechanism being tested uses the query to extract possible answers, which the notional user can then judge for relevance. But this cult of the perfect user is unwarranted—the relationship of the user to the system depends on the kind of information need. A journalist seeking background may want one relevant document; lawyers tend to require all relevant documents; students exploring a topic may only be interested in typical documents; and researchers are only interested in papers that agree with their point of view.

These instances of user kinds can be classified according to certain types of behaviour. These kinds can be illustrated as follows. First there is Pip, the impatient. Pip wants immediate gratification, and turns off the computer as soon as he finds what he wants, often necessitating the use of "scandisk". It is Pip that is being modelled by the common recall-precision approach to evaluating retrieval systems. Next there is Odysseus, the dogged, who will search to the bitter end and leave no document unread. Ismael, the exploratory, wants to understand the nature of his quest and is typical of people who might use a hypertext system or the tangent measure. Hamlet, the confused, doesn't know what he wants.[1] And last is Ophelia,[2] the deranged, whose queries don't make any sense. Thus the meaning of relevance judgements must be with regard to the kind of user being modelled.

Relevance judgements must also be considered in the context of the era in which they were made. The growth of information retrieval corresponded with the popularity of Sartre and existentialism, so that answers simply were; their meaning and content was not relevant [19]. Later, the influence of popular mysticism and Zen philosophies led to a reverse approach, in which answers were not [20]. Anarchists insisted that answers be statements that undermine the

[1] The problem of confusion should not be underestimated. It is remarkable how often one of us (S.W.) is confused with his namesake, Mr. Whiskers, despite the fact that the latter is a television personality with a different name.

[2] The paucity of female user types is due to the persistent stereotyping of women in Western literature, not to any bias on the part of the author.

question. Another approach has been to attack the implicit dominance of the query and ask whether the query is relevant to the answers [21], thus seeking equality in the query-answer relationship. That is, in this approach relevance is dependent on consensus and mutual understanding.

Politicisation is another factor. Answers are only correct in the contemporary political and opinionative climate. Some words, such as "environment", acquire new meanings over time, but relevance can shift even where word meanings do not. For example consider the following text.

> There was a collective God-help-us, a liturgical shriek of anguished relief, that rose at one moment from the assembled faithful. No surprise to the makers, but still! an achievement in all eyes. Sparks drummed and switched down the wires and coils—it was electric, but it was alive. [22]

This text would not previously have been judged relevant to a query on "artificial intelligence", but in the future, who knows?

Imagine yourself living in a stable world, in which documents are information and scientific papers such as this are knowledge. But nothing could be further from the truth—even the foundations on which theory can be based are uncertain. It is possible that knowledge and method may be universal at an undiscovered meta-level, an exciting prospect for future research into meaning. In the meanwhile, the way forward is not to grapple with the elusive smoke of similarity computation. The solution is empowerment of the user, engaging the user with the system such that distillation and refinement of the user's interactions are all brought to the benefit of the system, and only measuring the system in the context of a user, not in a potentially meaningless abstract case. Combining various new strategies it is shown below how an interface can allow easily achieved high effectiveness.

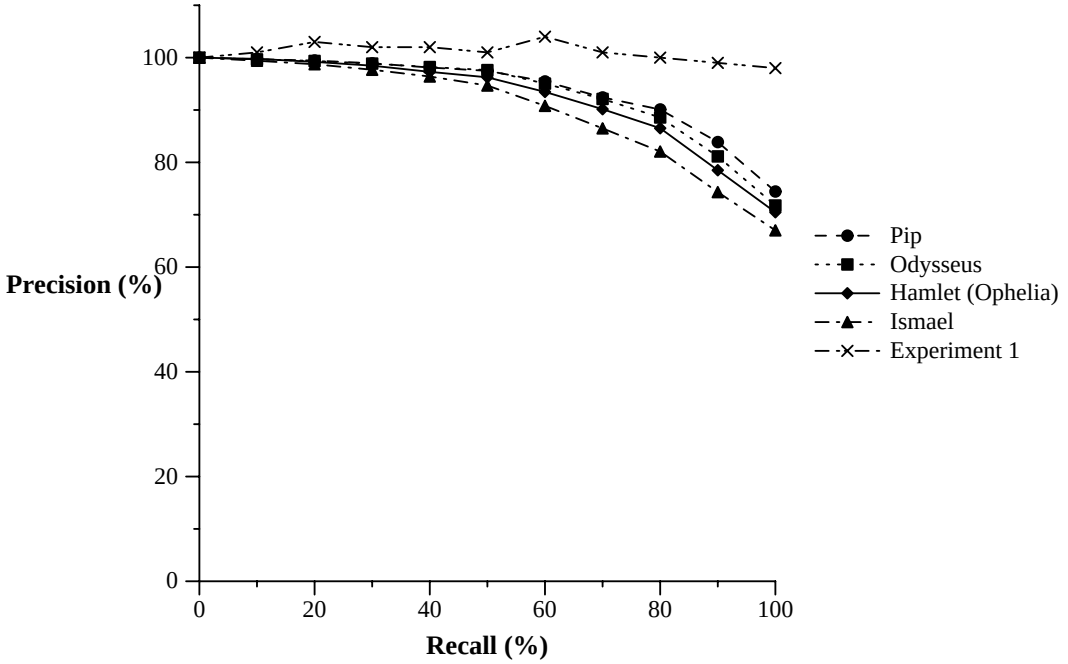
4 Experiments

Many experiments rely on a standard test collection. However, the intention was to measure users and therefore it is important to choose data over which prior knowledge could be asserted, wishing to not be constrained by what could be judged to be prior failures. This was important, since it was important to not have any variables in the systems: effectiveness measurements should be controlled. The data used was a collection of the author's own manuscripts, such as this paper, integrated with published material to provide balance. It was ensured that each user had not seen any of the documents in the collection prior to the experiments.

The interface was also kept constant, but is based on an important principle, warmth and friendliness, and is designed to minimise interaction costs [23]. The system used a popular world-wide web interface—called KROID—augmented with features such as animated images, contrasting colours, rapidly blinking text, and other techniques to highlight important aspects of the interface for usability. Users were so satisfied with answers through the interface that—by and large—after using the tools briefly, users had information needs satisfied and did not return frequently.

Having fixed each variable that is conventionally varied in other experiments, the user's approach

to information retrieval was investigated. Several approaches to such modifications were investigated, where each incorporates a strong process of reward for success and different querying tasks. These important variations are described below.



In the first experiments, users were rewarded for successfully identifying known relevant documents from the collection in response to queries [24]. This experiment varied and added enthusiasm, since users are often lazy and do not try hard enough to make the best use of the system [25]. The user was encouraged to use the interface for as long as required to produce a ranked set of KROID results, where each submitted final run was assessed financially. Several users were selected from the student population at Endeavour Research and Development (BVI), where each student was an international student from a taxed background. It was communicated to each user how to assess relevance [23]. This experiment showed that students could find nothing but relevant documents when it was explained that this was a factor in successful renumeration of the experiment.

The second experiments investigated the common user failing of missed relevance. In this experiment users assessed each answer using strict relevance assessment criteria that were provided. In brief, this criterium standardised the assessment process by asking simple questions that define relevance clearly to the user by assessing relevant documents as those that contained new information that was retrieved in response to query processes as created. Careful subjective assessment ensured that in this experiment effectiveness could reach at least 100%.

The third experiments were a turning point in research and a real contribution. In these experiments, the emotional component of a search was amplified, immersing the user in a total information retrieval experience. Searching can be frustrating when answers are irrelevant. Searching can be rewarding when responses meet an information need. For answers that are partially relevant, the process can be laborious or just plain boring. To improve the system beyond belief, detection of a relevant answer was rewarded with the ringing of a cheerful bell and re-amplification of the negated responses decribed below. For unsuccessful searches,

environmental factors were negatively modified: the response of the spherical module of the pointer device was randomly altered, the tactile response of the data entry interface was limited, and the visual interface was dimmed or spiked. For partially relevant searches, a hybrid approach was used: instead of negating environmental performance or restoring such performance, quasi-environmental factors such as colour vibrancy, MIDI sound scaping, and the degree of interface animation were moderated. Environmental modification works well—such techniques have been introduced in the student laboratories and (because the under-trained students are not adept at information retrieval) the laboratory is clearly affected by a negated environment. Clearly, better querying will improve the environment.

On a technical note, the project was made feasible by a new access technique of our invention, called FIND. FIND allows the direct comparison of all stored data with the query expression in very little time, thus providing rapid retrieval. The implementation in highly structured code was of high performance on our recently acquired Pentium computer system with the Voodoo 3D graphics card and 16-bit sound processing, running Windows NT 4 service pack 3 (install date 12 October 1998).

5 Conclusions

Much of the research in Information Retrieval has concerned improvements to similarity computations, statistics gathering, and term extraction, with a goal to improve effectiveness. However, a simple examination of user characteristics can readily show, the method of computing similarity is less important than the behaviour of the system interface and environmental factors. It was hypothesised there must be knowledge of the relationship between a query, its user, the environment, and the query and user instantiation in the real world! This hypothesis and others are demonstrated. With facilities for interaction and feedback appropriately incorporated, effectiveness of 100% can be achieved.

6 Acknowledgement

I would like to thank my supervisor, Joel Stubinz, for his helpful comments and help on this work.

References

[1] W.B. Frakes and R. Baeza-Yates, editors. *Information Retrieval: Data Structures and Algorithms*. Prentice-Hall, 1992.

[2] G. Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, Massachusetts, 1989.

[3] G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.

[4] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, second edition, 1979.

[5] S. Johnson. *The English Language*. Press Academie, London, 1981.

[6] C.T. Onions, editor. *The Shorter Oxford Dictionary on Historical Principles*. Oxford University Press, New York, 1973.

[7] F.U. Enyham. Jargon into slang and other transmogrifications. *J. Lang.*, pages 812–831, 1977.

[8] C. Darwin. *The Voyage of the Beagle*. Penguin, Cambridge, 1843.

[9] J.C. Maxwell. Equations. *J. Phys.*, pages 1–33, 1864.

[10] S. Strachan. *Ego is not a dirty word*. Ballantine, New York, 1988.

[11] Z. Belstin and J. Boltzis. Adrift on a sea of metaphor. In *Proc. Linguistica*, pages 212–219, Avignon, France, 1991.

[12] K. Krensen. *Real Time For Meaning: Smiley*. KK Publishing, Annandale, Sydney, 1992.

[13] T.B. Callow. A theory of the meaning of intent. *J. Semantics*, pages 1018–1094, 1969.

[14] B. Ashington. An automatic system for retrieval of electronic documents. In *Third British Colloquium on Electronic Computing*, pages 118–121, Manchester, 1956.

[15] G. Ackroyd. Finding documents: New approaches based on statistical methods. *J. Lang.*, pages 212–253, 1968.

[16] B. Ravino and R. Boll. A natural semantics for information retrieval. In *East Pacific Rim Symposium on Applied Linguistics*, pages 161–163, 1993.

[17] F. Haygull and W.G. Stein. Diachronics for identification of high-valency document interrelationships. *Int. J. Pr. L. Sys.*, pages 81–99, 1990.

[18] D. Carte. *Users and dualism*. Press Press, New York, New York, 1991.

[19] J. Kerrack. *A discourse on meaning*. Penguin, London, 1953.

[20] E.Y. Musashi. *Shadows of meaning: A personal voyage to enlightenment*. Porto, New York, 1966.

[21] B. Tompkins. Holistic computation. *Int. J. Comp. Rep.*, pages 191–207, 1986.

[22] B. Lytton. Machines and man. *Int. J. Comp. Rep.*, pages 18–21, 1983.

[23] C.E. Shannon. A mathematical theory of communications. *The Bell Systems Technical Journal*, 27:379–423, 1948.

[24] I.H. Witten, A. Moffat, and T.C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Van Nostrand Reinhold, New York, 1994.

[25] G.K. Zipf. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley, Reading, Mass., 1949.

[26] D.E. Knuth. *The Art of Computer Programming, Volume 3: Sorting and Searching, Second Edition*. Addison-Wesley, Massachusetts, 1973.