Some Notes on the Economics and Evaluation of Automatic Retrieval and Filtering of Communication Goods

Richard Bergmair

November 2, 2007

Abstract

In this paper, we address the problem of searching for or filtering electronic documents in large collections such as the world wide web. This will be framed as an economic resource allocation problem, while invoking automatic retrieval engines as a non-market pricing mechanism.

For information retrieval, this point of view helps illuminate many imporant questions, such as: What is the benefit of a relevant document being retrieved? What is the impact of an error that results in an irrelevant document being retrieved? What is the impact of an error that results in a relevant document not being retrieved? How can these be measured, compared, and weighted against each other?

For economics, this point of view shows how retrieval engines can be invoked to mitigate many of the weaknesses of traditional markets when it comes to dealing with information and attention in the context of communication. How can we reduce search costs, given the vast catalogue of communication goods on offer in real-world communication economies such as the internet? How can we invoke competition, given the highly differentiated nature of communication goods such as newsfeeds, newsgroups, or websites? How can we deal with the fact that communication goods are experience goods?

This paper is theoretic in nature, and interdisciplinary in its approach.

"[...] in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it [...]" (Simon, 1971)

1 Motivation & Introduction

Herein, we shall have a few comments to make about the relation between the fundamentals of economics and information retrieval. As a point of departure, it is perhaps interesting to note that the two fields have quite substantial overlap, in that they are both studying the efficient allocation of scarce resources among competing ways of putting them to use, e.g. the efficient allocation of flats in central London among possible tenants or the efficient allocation of users' attention to documents on the web. When it comes to the automatic allocation of electronic documents and the attention of potential readers, both seen as economically valuable goods, these areas of study have important insights to offer each other. In the rest of this paper, we will collectively refer to these goods as communication goods.

This paper is theoretic in nature and interdisciplinary in its approach. To the extent possible, we will try to start from first principles, in order to make all the material reasonably accessible to readers from either an information retrieval or economics background. Readers may wish to consult the textbook by Baeza-Yates and Ribeiro-Neto (1999) on information retrieval, the book by Voorhees and Harman (2005) on retrieval evaluation, the textbook by Varian (2006) on economics, or the book by Shapiro and Varian (1998) on the economics of information goods.

Communication economies, as they arise for example on the web, for email, and for newsgroups and newsfeeds, need to be efficient in the sense that they should yield an allocation of information and attention, such that every possible reallocation would be opposed by some participant. In real world communication economies, such inefficiencies are currently commonplace, email spam being the most obvious kind. This is due to the fact that communication economies lack market mechanisms of the kind assumed by economic theory: goods do not have perfect substitutes; search costs are potentially prohibitive; and goods have to be consumed for their quality to be assessed (i.e. the experience good problem). Current search engines and document filters offer a resource allocation mechanism alternative to the markets traditionally studied by economists. They invoke a notion of substitutability by which two given communication goods are substitutes, if they are equally relevant to a given information need, for example as expressed by a search query. Since they are fully automated and operate on a large scale in a centralized fashion, they run up minimal search costs. Finally, they act as recommender systems, offering a way around the experience good problem.

In this paper, we will study the resource allocation problem that arises for communication goods, both from the point of view of economics and information retrieval, by developing a mathematical model that has a dual interpretation. On one hand, it serves the economist as a characterisation of the economic function of a retrieval engine as a non-market price-setting mechanism. On the other hand, it serves the researcher in information retrieval as an evaluation mechanism measuring the economic efficiency of a given search strategy.

2 Related Problems

In this section, we will quickly review some open problems raised so far in the literature on information retrieval and economics. The rest of this paper will be devoted to laying some theoretical groundwork for resolving those problems in an interdisciplinary manner.

Information Retrieval Problems

(i) Cooper (1973) suggested the use of cardinal document utilities, in the sense of gradual degrees of relevance for any given document along a polyvalent scale rather than a simple dichotomy. Nowadays, such document utilities are important for web retrieval engines. – But how can we interpret, and perhaps even measure, such general document utilities?

(ii) Traditional evaluation measures like precision and recall are based on dichotomous relevance judgements and cannot straightforwardly be applied to the more general document utilities mentioned before. – How can we generalise existing evaluation measures, or devise new ones, for use with general document utilities?

(iii) Evaluation methods that directly compare precision and recall, such as F_β (van Rijsbergen, 1979), precision-recall graphs, precision-recall break-even points, etc., are now ubiquitous. These rely on fixed weightings β to determine just how much precision can be traded off for a given unit of recall. – But there is currently little theoretic basis upon which to interpret and optimize precision-recall tradeoffs β for given applications.

(iv) For ranked retrieval, evaluation measures rely on cutoffs along document rankings. TREC (Voorhees and Harman, 2005) reports both precision and recall for cutoffs of 5, 15, 30, 100, and 200. – But there is little consensus about which number to look at and optimize.

(v) For purposes of comparative system evaluation, results of individual retrieval runs need to be aggregated across different users and information needs with different numbers of relevant documents. This is commonly done by micro- or macroaveraging the usual precision-recall-based measures. – But problems with these modes of aggregation have been identified early on (Robertson, 1969) and are still a pressing concern at TREC (Buckley and Voorhees, 2005).

(vi) Certain information management problems adjacent to the core retrieval ranking problem are well known to have a significant impact on the user experience presented by a given system. For example, the user might prefer a shorter document to a longer one when both are equally relevant. Similarly, a user might want the system to identify, highlight, and summarize relevant passages in long documents.

– It is currently not well understood how performance on these features should be weighted and how they relate to each other.

Economic Problems

(vii) Attention is a scarce resource. Recently, the notion of the attention economy has been suggested in the popular literature (Davenport and Beck, 2002; Lanham, 2006; Franck, 2007). In such an attention economy, we take the availability of information at no cost for granted, which leaves as the core economic problem the efficient allocation of attention. – For communication goods, however, scarce resources like labour have to be expended to make available information.

(viii) The production of information goods as studied by Varian (1995, 2000) and others generates high fixed costs for providing the first unit, and close to zero marginal costs for reproducing and distributing subsequent units. – The communication goods we have in mind, however, internalize the attention expended by consumers. These marginal costs should not be generally neglected.

(ix) On perfectly competitive markets, incumbents and possible market entrants are able to profitably produce perfect substitutes for overpriced goods. – However, neither the information contained in a particular information good, nor the attention devoted to it by a particular person, have perfect substitutes.

(x) Traditional search mechanisms (Nelson, 1970) would involve consumers manually assessing the quality and prices of all goods on offer. – This would obviously run up search costs to prohibit any kind of trade, if the catalogue of goods to be searched was, for example, the web with its vast number of communication goods competing for attention.

(xi) Communication goods are experience goods (Varian, 1995; Nelson, 1970), i.e. their value only becomes apparent after they have been consumed. This issue has usually been dealt with in the economics literature by assuming that mechanisms like preview, review, or reputation give the consumer prior knowledge about the value of a good, thus essentially negating the hypothesis. – For communication goods like emails, these mechanisms are not applicable.

3 Prices and Set Retrieval

Let us first introduce some notation and terminology. We will denote a document in a given collection or the communication good it represents, i.e. the information it contains and the attention devoted to reading it, by the variable d. Then $\mathrm{P}(d)$ denotes the Cooper-utility[1] (Cooper, 1973) or reservation price of d as assigned by a user or consumer. An

[1] It is important to note that Cooper's notion of utility is that of a cardinal utility. This terminology is somewhat unfortunate for our exposition. Therefore it is important to keep in mind that "utility" as a technical term in retrieval corresponds, in our model, to the economic notion of "reservation price", not to the economic notion of "utility". We therefore speak of "Cooper-utilities", where we refer to the utility notion usually used in information retrieval.

associated technical term in the retrieval literature is that of relevance. We say d is relevant iff $\mathrm{P}(d) > 0$ and irrelevant otherwise. We write $\hat{\mathrm{P}}(d)$ to denote d's retrieval status value as assigned by a retrieval engine, or d's price as set by the non-market pricing mechanism which it embodies.

Taking the point of view of economics, let us, in a first step, interpret the retrieval engine which maximizes user satisfaction as measured by the number of documents both relevant and retrieved as the pricing mechanism which maximizes the volume of trades as measured by the total amount of money that changes hands.

For every transaction between a given producer and a given consumer, there is a transaction value $\mathrm{V}(d)$ associated with each good d on offer, which is ultimately realized as a marginal revenue for the producer and a marginal cost for the consumer. This is

$$\mathrm{V}\big(\,\hat{\mathrm{P}}(d),\ \mathrm{P}(d)\,\big) = \begin{cases} \hat{\mathrm{P}}(d) & \text{if } \hat{\mathrm{P}}(d) \leq \mathrm{P}(d), \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Here the assumption is that the consumer evaluates the offer by ascertaining a reservation price $\mathrm{P}(d)$. In the case $\hat{\mathrm{P}}(d) \leq \mathrm{P}(d)$, where the reservation price exceeds the offer price, the good is transacted for the offer price $\hat{\mathrm{P}}(d)$. Otherwise the good is not transacted, so the transaction value is zero.

The ideal pricing mechanism will set prices in such a way that (1) $\forall d : \hat{\mathrm{P}}(d) \leq \mathrm{P}(d)$, because otherwise the number of goods that change hands could be increased by reducing the prices of goods which the consumer finds too expensive. (2) $\forall d : \hat{\mathrm{P}}(d) \geq \mathrm{P}(d)$, because otherwise the amount of money that changes hands could be increased by increasing the prices of goods which the consumer would still be willing to buy if they were more expensive. This leads to the following two evaluation measures:

$$\textsc{Prec} = \frac{\sum_d \mathrm{V}\big(\,\hat{\mathrm{P}}(d),\ \mathrm{P}(d)\,\big)}{\sum_d \hat{\mathrm{P}}(d)}, \text{ and} \tag{1}$$

$$\textsc{Rec} = \frac{\sum_d \mathrm{V}\big(\,\hat{\mathrm{P}}(d),\ \mathrm{P}(d)\,\big)}{\sum_d \mathrm{P}(d)}. \tag{2}$$

We can see $\sum_d \mathrm{V}(\cdots)$ as the producer's revenue or the consumer's costs associated with the consumption bundle chosen. This is evaluated, in the case of $\textsc{Prec}$, as a proportion of the value we would have observed if prices were low enough so that the consumer would buy all goods or, in the case of $\textsc{Rec}$, if prices were high enough so that every good bought by the consumer would change hands for its reservation price. We have $\textsc{Prec} = 1$ iff condition (1) is fulfilled, and $\textsc{Rec} = 1$ iff condition (2) is fulfilled. They are both one iff $\mathrm{P} = \hat{\mathrm{P}}$, and they both decrease when $\sum_d \mathrm{V}(\cdots)$ decreases.

Given these definitions, as motivated from an economic point of view, it is now easy to see how the evaluation measures traditionally used in information retrieval relate to our model. To make this formally explicit, we need to define a quantisation operator $\mathbb{1}\{\cdot\}$ as follows:

$$\mathbb{1}\{x\} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x > 0, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

The traditional evaluation measures are defined in terms of the bivalent set $\textsc{Ret} = \{d|\hat{\mathrm{P}}(d) > 0\}$ of documents retrieved by a system and the bivalent set $\textsc{Rel} = \{d|\mathrm{P}(d) > 0\}$ of relevant documents. For the ideal retrieval system (1′) $\textsc{Ret} \subseteq \textsc{Rel}$, so that the user never gets to see an irrelevant document, and (2′) $\textsc{Ret} \supseteq \textsc{Rel}$, so that the user gets to see all relevant documents. – This leads to the set overlap measures of precision and recall, which can now be seen as quantized versions of our earlier definitions:

$$\textsc{Prec}' = \frac{|\textsc{Ret} \cap \textsc{Rel}|}{|\textsc{Ret}|} = \frac{\sum_d \mathrm{V}\big(\ \mathbb{1}\{\hat{\mathrm{P}}(d)\},\ \mathbb{1}\{\mathrm{P}(d)\}\ \big)}{\sum_d \mathbb{1}\{\hat{\mathrm{P}}(d)\}}, \tag{1′}$$

$$\textsc{Rec}' = \frac{|\textsc{Ret} \cap \textsc{Rel}|}{|\textsc{Ret}|} = \frac{\sum_d \mathrm{V}\big(\ \mathbb{1}\{\hat{\mathrm{P}}(d)\},\ \mathbb{1}\{\mathrm{P}(d)\}\ \big)}{\sum_d \mathbb{1}\{\mathrm{P}(d)\}}. \tag{2′}$$

Note that $\textsc{Prec}' = 1$ iff condition (1′) is fulfilled, and that $\textsc{Rec}' = 1$ iff condition (2′) is fulfilled. They are both 1 iff $\textsc{Ret} = \textsc{Rel}$, and they both decrease with decreasing $|\textsc{Ret} \cap \textsc{Rel}|$, where $\textsc{Prec}'$ measures the error as a proportion of $|\textsc{Ret}|$, and $\textsc{Rec}'$ measures it as a proportion of $|\textsc{Rel}|$. From the point of view of information retrieval, these properties intuitively justify the use of these measures for evaluation purposes.

The reader can verify that (1) and (2) always imply the weaker conditions (1′) and (2′) respectively. In information retrieval, the assumption is often made, for experimental purposes, that $\mathrm{P}(d) \in \{0, 1\}$, so that documents are a priori either relevant, and have a reservation price of one dollar, or irrelevant, in which case they have a reservation price of zero dollars. Furthermore, one could assume that the system engages in proper set retrieval with $\hat{\mathrm{P}}(d) \in \{0, 1\}$, rather than ranked retrieval, which means that documents are either retrieved, so that our pricing scheme sets a price of one dollar, or not retrieved, so that it sets a price of zero dollars. In this special case, the two definitions are in fact strictly equivalent, as the quantisation becomes entirely vacuous.

Notes on Economics

Except for the effects of quantisation in experimental retrieval evaluation, the goal of information retrieval, maximizing precision and recall, can therefore be understood, under our model, in economic terms as the goal of the monopolistic producer of information goods considered by Varian (1995, 1996, 2000), which is to set prices equal to the consumers' reservation prices. A retrieval engine can then be understood as a price discrimination mechanism, very similar in its effects to the mechanisms usually considered for information goods, such as product discrimination and versioning (Varian, 1997) or bundling (Bakos and Brynjolfsson, 1999). However, we believe that retrieval engines have many advantages, as they can potentially act as first-degree price discriminators.

Concerning Varian's model of information goods, it is important to note that this primarily applies to goods such as packaged software, music CDs, home videos, computer games etc. We will, on the other hand, be interested in a class of goods we shall refer to as communication goods. These are e-mails, or webpages, for example, and do not give rise to the same kind of natural monopoly that arises for traditional information goods. Similarly to information goods, one might argue that the information content of webpages or e-mails is highly differentiated, so that producers do not need to compete on the provision of information. However, treating these goods as traditional information goods would be neglecting the fact that producers of communication goods have to compete for the attention of their consumers. In the next section, we will outline an alternative to or extension of Varian's model which covers these phenomena.

Notes on Retrieval

Our economic variant of the traditional evaluation measures has the nice property that it is well defined and readily interpretable also for reservation prices and offer prices, i.e. relevance judgements and retrieval status values, defined anywhere along a polyvalent cardinal scale, not just for dichotomies.

This, of course, is the problem we mentioned earlier under point (i), and which has first been raised by Cooper (1973) in his "naive evaluation methodology". The notion of utility he introduced to the field of information retrieval can be likened to that used in utilitarianist economics. Here utilities are measured along a cardinal scale, and they are generally seen as universal in some sense. He uses utility as "a cover term for whatever the user finds to be of value about the system output, whether its usefulness, its entertainment or aesthetic value, or anything else". He considers a thought experiment in which he determines the utilities of documents by asking users how much money $\mathrm{P}(d)$ they were willing to give up in order to get document d. This concept is, in modern economics, that of a reservation price, while economists now usually think of a consumer's utility function as ordinal.

Taking this naive evaluation methodology as a theoretical point of departure, Cooper then proposes an "implementation of the philosophy". Here he makes a concession to the practical feasibility of experiments in assuming that utilities are not observed along a polyvalent cardinal scale, but only in the form of dichotomous relevance judgements. Similarly, Robertson employs a notion of unobservable continuous "synthema" (Robertson, 1976, 1977a) which underlies observable dichotomous relevance judgements. He refers to Cook's threshold model of relevance (Cook, 1975), which offers one way to view the relation of relevance as a continuum and relevance as a dichotomy. This seems to be a sound experimental procedure for comparative system evaluation because, although relevance judgements are dependent on scale form (Katter, 1968), rankings of systems based on evaluation measures such as precision and recall are surprisingly robust to varying relevance judgements (Lesk, 1969; Buckley and Voorhees, 2005).

Once we can measure $\hat{\mathrm{P}}(d)$ in terms of general document utility, it only seems natural to apply it as a ranking criterion, maximizing the expected Cooper-utility at every cutoff. This could be seen as a generalisation of the probability ranking principle (Robertson, 1977b, 1997), as the probability of a given document to have unit utility can always be seen as the expected value for the utility when only unit and zero utilities are permitted (Cooper and Maron, 1978; Cooper, 1978).

There is now a new need to understand the ranking criterion employed by a retrieval engine as a polyvalent scale rather than just a dichotomy. Web search engines apply continuous measures for the a-priori quality of a web document. For example PageRank (Page, 1998; Brin and Page, 1998), as a factor determining a Cooper-utility, is in direct violation of Cooper's "assumption II". So it would not be justified under Cooper's model to break down PageRanks into dichotomous relevance judgements, and it would indeed seem to make little sense to do so.

In response to problem (ii), it is at this point natural to rank documents d by the values of $\hat{\mathrm{P}}(d)$ and to evaluate the resulting rankings by PREC and REC under definitions (1) and (2), i.e. without applying quantisations. This fulfills a number of important intuitions we have about evaluation schemes for ranked retrieval. Documents ranked higher have a greater potential impact on precision and recall than documents ranked lower, and documents which are more valuable to the user have a greater impact than documents that are less valuable.

What is more uncommon about this idea is the fact that retrieval status values are significant in terms of their cardinal values rather than just the rankings they impose on documents. We believe there is much to be gained and little to be lost from adopting this stronger evaluation criterion. There is little to be lost, as these evaluation measures respond to all errors that set-based measures would react to. On the other hand, they also react to marginal changes in retrieval status values, indicating whether the change had a positive or negative marginal effect. Robertson and Zaragoza (2007) point out that this kind of "smooth cost function" is advantageous for optimisation purposes, even if one is ultimately interested only in rank-based retrieval.

What might strike the reader as peculiar, though, is the fact that a ranking-preserving transformation can systematically be applied to retrieval status values in such a way as to affect these measures. Taking a given system and fixed reservation prices $\mathrm{P}(d)$ as a point of departure, one can positively affect precision and negatively affect recall by, for example, scaling all retrieval status values to smaller values. We will discuss this phenomenon in greater detail in the next section.

4 Costs and Relevance Thresholding

In a next step, we will assume the reservation price $\mathrm{P}(d)$ is composed of a component $\mathrm{R}(d)$ representing a unit return, and a component $\mathrm{C}(d)$ representing a unit cost. This models the fact that the sender of a given communication good needs to compensate the recipient for their attention, while the recipient needs to compensate the sender for the information. We will discuss this from an economic point of view in greater detail later in this section. For now let us simply set

$$\mathrm{P}(d) = \mathrm{R}(d) - \mathrm{C}(d).$$

Ideally, one would seek to separately quantify the $\mathrm{R}(d)$ and $\mathrm{C}(d)$ associated with each good d. This would yield the following evaluation measures under our earlier definitions.

$$\textsc{Prec} = \frac{\sum_d \mathrm{V}\big(\ \hat{\mathrm{P}}(d),\ \mathrm{R}(d) - \mathrm{C}(d)\ \big)}{\sum_d \hat{\mathrm{P}}(d)}, \text{ and} \tag{1}$$

$$\textsc{Rec} = \frac{\sum_d \mathrm{V}\big(\ \hat{\mathrm{P}}(d),\ \mathrm{R}(d) - \mathrm{C}(d)\ \big)}{\sum_d \mathrm{R}(d) - \mathrm{C}(d)}. \tag{2}$$

However there are two possible ways to idealise the model when these measurements are not possible. One possible simplification is to treat $\mathrm{C}(d) = c$ as a constant across all

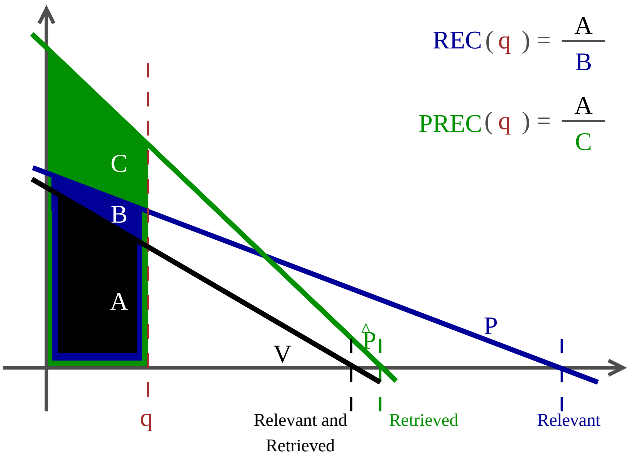

Figure 1: measures based on a consumption budget of q goods

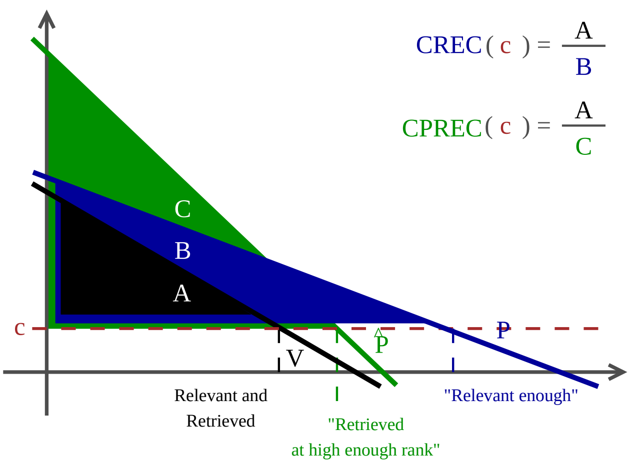

Figure 2: measures based on a constant cost of consumption c

goods d:

$$\text{CP}\textsc{rec}(c) = \frac{\sum_d \mathrm{V}\big(\, \hat{\mathrm{P}}(d),\ \mathrm{R}(d) - c \,\big)}{\sum_d \hat{\mathrm{P}}(d)}, \text{ and} \tag{3}$$

$$\text{CR}\textsc{ec}(c) = \frac{\sum_d \mathrm{V}\big(\, \hat{\mathrm{P}}(d),\ \mathrm{R}(d) - c \,\big)}{\sum_d \mathrm{R}(d) - c}. \tag{4}$$

The other possibility is to assume $\mathrm{C}(d) = 0$ for all d, while invoking a budget constraint q to limit the number of goods that can be consumed. This leads to the following evaluation measures:

$$\text{QP}\textsc{rec}(q) = \frac{\sum_{|\{d' \text{ s.t. } \hat{\mathrm{P}}(d') \le \hat{\mathrm{P}}(d)\}| \le q} \mathrm{V}\big(\, \hat{\mathrm{P}}(d),\ \mathrm{R}(d) \,\big)}{\sum_{|\{d' \text{ s.t. } \hat{\mathrm{P}}(d') \le \hat{\mathrm{P}}(d)\}| \le q} \hat{\mathrm{P}}(d)}, \tag{5}$$

$$\text{QR}\textsc{ec}(q) = \frac{\sum_{|\{d' \text{ s.t. } \hat{\mathrm{P}}(d') \le \hat{\mathrm{P}}(d)\}| \le q} \mathrm{V}\big(\, \hat{\mathrm{P}}(d),\ \mathrm{R}(d) \,\big)}{\sum_{|\{d' \text{ s.t. } \mathrm{R}(d') \le \mathrm{R}(d)\}| \le q} \mathrm{R}(d)}. \tag{6}$$

The relation between these different versions of the precision and recall measures is visualized in figures 1 and 2. The green triangle visualizes the prices or retrieval status values $\hat{\mathrm{P}}(d)$ assigned to every subsequent d, when these are sorted in descending order in the positive x-direction. Similarly, the blue triangle visualizes the return $\mathrm{R}(d)$ assigned to each d, when these are sorted by descending values of $\mathrm{R}(d)$. The black triangle visualizes the transaction values $\mathrm{V}\big(\, \hat{\mathrm{P}}(d),\ \mathrm{R}(d) \,\big)$ of each d, when these are sorted by descending values of $\hat{\mathrm{P}}(d)$. We can then apply rank-based cutoffs as in figure 1 or cost-based cutoffs as in figure 2. Recall and precision are one at a given cutoff when the blue and the green area respectively are the same magnitude as the black area. When this occurs at all cutoffs, the triangles are equal.

Notes on Economics

The model underlying markets for traditional information goods of the kind we discussed in the previous section assumes that information is in fact the sole resource being transacted. This is the case for example for packaged software, music CDs, home videos, computer games, etc. Recently there have been a number of publications centered around the notion of an attention economy (Davenport and Beck, 2002; Lanham, 2006; Franck, 2007). On markets for attention goods, the sole resource being transacted is attention. This is the case for example for advertising.

Our notion of a communication good, in contrast to a pure information good or a pure attention good, broadly follows the idea of Simon (1971) quoted earlier, which is that information can only be consumed if, at the same time, attention is consumed. A sender acts as a producer of information and consumer of attention at the same time, and a recipient acts as a producer of attention and consumer of information. A communication good is transacted as follows: Initially the sender produces an information good by expending scarce production factors such as labour. The recipient starts with an initial allocation of attention, which we can view as a kind of scarce natural resource. A communication good d can now be transacted if the information good is reproduced and transmitted by the sender for the recipient and an according quantity of attention is expended by the recipient.

There are two ways to analyze this scenario under the model outlined earlier in this section. One can either take the sender's point of view, where the price of information $P_I(d)$ is a unit cost $C(d) = P_I(d)$, while the price of attention is a unit return $R(d) = P_A(d)$; or, alternatively, one can take the recipient's point of view, where it is the other way around, i.e. $R(d) = P_I(d)$, and $C(d) = P_A(d)$. Since the two viewpoints are in fact symmetric, we will w.l.o.g. adopt this latter viewpoint. This means that positive prices on communication goods indicate a monetary payment by the recipient to the sender, while negative prices and transaction values would indicate a payment from the sender to the recipient.

For example, if d is a scientific article that is downloaded by an interested researcher over the web, we will have $R(d) > C(d)$. Such an article will therefore be transacted at a positive price $\hat{P}(d) > 0$, i.e. the publisher will be able to demand monetary compensation for its services. If, on the other hand, d is an e-mail that has found its own way into a mailbox and it contains no information of value to the recipient, we will have $R(d) < C(d)$. Such a spam mail will therefore be transacted at a negative price $\hat{P}(d) < 0$, i.e. the owner of the victimized mailbox should be able to demand monetary compensation.

There have been some suggestions in the literature concerning financial instruments to make possible this kind of bidirectional compensation, such as attention bonds (Loder et al., 2004) or interrupt rights (Fahlman, 2002). In practice, however, such instruments are not currently available in real communication economies such as the web. In this case one would restrict prices and transaction values to be nonnegative.

Therefore, a strategy that is often pursued in practice is to bundle communication goods in such a way that the resulting bundle d has $R(d) \geq C(d)$, because such a bundle can be transacted between sender and recipient at a nonnegative price $\hat{P}(d) \geq 0$. For example, a webpage will often contain some valuable primary content and some advertising. Web users are compensated for enduring worthless advertising by also receiving valuable content. Content providers are compensated for their labour by monetary payments from advertisers. Advertisers are compensated for their monetary expenses by getting the attention of consumers. The same model is used for free-tv, with the tv-station acting as an additional intermediary.

A few words are in place about how our notion of a communication good, which integrates previous ideas about information goods and attention goods, extends on the account given in the literature so far.

As noted under (vii), the notion of an "attention economy", as coined in the popular literature (Davenport and Beck, 2002; Lanham, 2006; Franck, 2007), in its strongest formulation, implies that units of attention replace money as we know it. The same "new economy" type of claim has, depending on the issues of the day, also been made about CO_2 or oil. Our model certainly implies nothing quite as grandiose. However it does recognize the need to trade attention for other scarce resouces like CO_2, oil, or even the less fashionable ones like services, labour, consumer goods, or land. One means by which such trade can be achieved is, of course, precisely money as we know it, and our model can be invoked in that way.

Furthermore, we believe it is quite incorrect to assume, as advocates of the attention economy sometimes do, that information is not a scarce resource in any sense, simply because information can be reproduced and distributed cheaply. It is certainly conceivable how this false impression might arise for information consumers perceiving the world through the web. But nothing could be further from the truth for information producers whose job it is to make information available on the web by expending their labour and other scarce production factors. This is why, under our model, the social cost of the information available on communication markes is not necessarily zero, thus providing a possible solution to what we have introduced as problem (vii) before.

Concerning previous models of information goods, it is important to note that, by internalizing the cost of attention in the profit maximisation problem of a producer of communication goods, the marginal costs faced by the producer of a communication good are not zero, as would be the case for information goods of the kind studied by Varian (1995, 2000) and others. We have pointed this out under (viii). This leads Varian to conclude that markets for information goods are natural monopolies. For the communication goods we have in mind, this monopoly price is still valid as an upper limit for any efficient price. There might, however, be lower efficient prices, due to the fact that a consumer's attention might, in itself, be worth something to the producer. Intuitively, producers of communication goods will have to compete on the attention of consumers, when the relative price of attention is high compared to the price of information.

Given this conception of a communication good, the evaluation measures introduced earlier in this section take on an economic interpretation.

For CPREC(c) and CREC(c), we assume that communication goods are "packaged" in such a way, that the attention they consume raises a constant unit cost c. Furthermore, we assume that consumers have perfect knowledge about the value of the information contained in a given communication good, without having to expend any search costs or costs of attention.

For QPREC(q) and QREC(q), we go even further in assuming that the costs of attention associated with any communication good are in fact zero. However, we relax the assumption concerning perfect knowledge of the value of all goods. Instead, we assume that a consumer, in a first step, uses price signalling to select the q goods with the highest prices. Only in a second step, does the consumer go on to obtain knowledge about the value of the information conveyed in those q goods.

Under both sets of assumptions, the resulting numbers can be interpreted, as before, as a proportion of social surplus extracted by a given pricing scheme, thereby representing the pricing strategy of a monopolist. In the next section, we will develop a more general account that takes into account the non-monopoly case and search costs.

Notes on Retrieval

The fact that the attention of users is limited also has a noticeable impact on the design of retrieval engines. For example, the implicit assumption made for simple set retrieval that a user will want to, and be able to, see all relevant documents, irrespective of how many there are in total, is usually relaxed for ranked retrieval. The user is presented a ranking of documents and is expected to examine only the top q. This model is accommodated by the following traditional evaluation measures:

$$\text{QP}{\small\text{REC}}'(q) = \frac{\sum_{|\{d' \text{ s.t. } \hat{\text{P}}(d') \leq \hat{\text{P}}(d)\}| \leq q} \text{V}\big(\, \mathbb{1}\{\hat{\text{P}}(d)\},\ \mathbb{1}\{\text{R}(d)\}\,\big)}{q}, \tag{5'}$$

$$\text{QR}{\small\text{EC}}'(q) = \frac{\sum_{|\{d' \text{ s.t. } \hat{\text{P}}(d') \leq \hat{\text{P}}(d)\}| \leq q} \text{V}\big(\, \mathbb{1}\{\hat{\text{P}}(d)\},\ \mathbb{1}\{\text{R}(d)\}\,\big)}{\sum_d \mathbb{1}\{\text{R}(d)\}}. \tag{6'}$$

We have already seen why QPREC and QREC lend themselves to an economic interpretation. In the following we will discuss their interpretation in an information retrieval context, showing some advantages of these definitions over the traditional measures QPREC′ and QREC′.

First, note that our QPREC and QREC measures are different from the traditional measures QPREC′ and QREC′ in that they remove quantisations and allow general Cooper-utilities rather than being restricted to dichotomous relevance judgements. We believe that the quantisations here are counterproductive as they artificially enforce all ranking decisions up to rank q to have equal weight, while our measures would give greater weight to higher ranks.

Secondly, our normalisations seem to be better behaved. For example, assume there are only two relevant documents. QPREC′(10) would, in this case, be indifferent at QPREC′(10) $= 0.2$ between the perfect system that retrieves exactly the two relevant documents and a system that also retrieves eight irrelevant documents and scatters the two anywhere among the ten. QPREC(10), on the other hand, would reward the first system for knowing that there are only two relevant documents by giving it the full score QPREC(10) $= 1.0$.

As another example, assume there are fifty relevant documents. QREC′(2) is now indifferent at QREC′(2) $= 0.2$ between a perfect system that retrieves exactly the most relevant and second most relevant document and a system that retrieves the 49th and 50th most relevant document. QREC, on the other hand, would reward the first system for knowing which documents are the two most relevant ones by giving it a perfect score QREC(10) $= 1.0$. This is, of course, not possible in a traditional setup when there is no notion of gradual relevance to begin with.

From a theoretic point of view it is quite advantageous that our numbers for QPREC and QREC always range from zero to one, while the range of QPREC′ and QREC′ depends on the cutoff chosen and the total number of relevant documents beyond the cutoff. This makes these numbers somewhat confusing to interpret and compare, and it has been pointed out on several occasions (Robertson, 1969; Buckley and Voorhees, 2005) that this leads to problems with macroaveraging. Our scheme therefore provides a possible approach to problem (v).

Thirdly, in response to problem (vi), our model would accommodate for the evaluation of summarization and similar components in an information retrieval system. For example, one could make $\text{C}(d)$ a function of the length of the documents returned, the intuition being that it costs less attention to consume a short document than a long one.

In the previous section, we mentioned the peculiar property of our scheme to assign a significance to the cardinal values of retrieval status values, rather than just the rankings they impose. This is a direct consequence of the definitions we made in this section about cost-based cutoffs – more particularly the fact that we want PREC $= 1$ and REC $= 1$ iff

$\text{CPREC}(c) = 1$ and $\text{CREC}(c) = 1$ for all c. If this is the case then, for a given c, a transformation of $\hat{\text{P}}(d)$ that systematically assigns smaller values, even if this is done in a ranking-preserving fashion, will mean that some d will not be retrieved at cutoff c that would otherwise have been retrieved, which would have a negative impact on $\text{CREC}(c)$, and hence REC.

5 Artificial Markets for Communication Goods

Our exposition so far has been structured along the lines of traditional retrieval evaluation. We have introduced different evaluation measures and given them a dual interpretation in economics. This has served to shed some light on relations between some very basic notions of information retrieval and economics, such as reservation price and relevance, or price and retrieval status value. We have seen that the goal of optimizing precision and recall in its various forms is equivalent to the goal of the monopolistic producer of communication goods.

In this section, we will make a paradigm shift. We will introduce the economic agents involved in retrieval, express their profit functions on the basis of a given non-market pricing scheme, and consider their profit maximization problems. Finally, we will be able to quantify the efficiency of the resource allocation achieved. This notion of economic efficiency can then serve as a possible new retrieval evaluation measure.

First, consider the producer of d. The producer makes a production decision denoted by the predicate $\text{produce}(d, \hat{\text{P}})$ which is true iff the producer decides to produce good d, given a set of prices $\hat{\text{P}}(\cdot, d)$. For each good d produced, the producer faces a fixed cost $\text{FC}_\text{I}(d)$ reflecting the scarce production factors that go into authoring, editing, and publishing the document. However, following Varian, we assume there are no unit costs directly associated with reproducing or distributing information. Furthermore, the producer makes a supply decision, denoted by the predicate $\text{supply}(c, d, \hat{\text{P}})$ which is true iff the producer is willing to supply good d to a specific consumer c for a specific price $\hat{\text{P}}(c, d)$. On each good d supplied and demanded, the producer faces a price $\hat{\text{P}}(c, d)$ as a revenue and a return $\text{R}_\text{A}(c, d)$ reflecting the value of the consumer's attention to the producer.

Next, consider consumer c. The consumer needs to conduct a search in order to determine the quality of any good d supplied by any producer and set a personal reservation price accordingly. So on each good d produced and supplied by some producer, the consumer faces a search cost $\text{C}_\text{S}(c, d)$ which reflects the amount of attention they have to expend in order to determine whether or not to demand d. We denote this demand decision as a predicate $\text{demand}(c, d, \hat{\text{P}})$ which is true iff the consumer is willing to buy good d at price $\hat{\text{P}}(c, d)$ Then, on each good demanded, the consumer faces the price $\hat{\text{P}}(c, d)$ as a cost, the cost of attention $\text{C}_\text{A}(c, d)$, and a return on information $\text{R}_\text{I}(c, d)$ reflecting the inherent value of the information conveyed.

We can now express the consumer's and the producer's total profit functions as:

$$\textsc{Tp}_{\mathrm{C}}(c) = \sum_{d|\mathrm{produce}(d,\hat{\mathrm{P}})\wedge\mathrm{supply}(c,d,\hat{\mathrm{P}})} -\mathrm{C}_{\mathrm{S}}(c,d) + \begin{cases} \mathrm{R}_{\mathrm{I}}(c,d) - \mathrm{C}_{\mathrm{A}}(c,d) - \hat{\mathrm{P}}(c,d) & \text{if demand}(c,d,\hat{\mathrm{P}}), \\ 0 & \text{otherwise;} \end{cases}$$

$$\textsc{Tp}_{\mathrm{P}}(d) = -\textsc{Fc}_{\mathrm{I}}(d) + \sum_{c|\mathrm{supply}(c,d,\hat{\mathrm{P}})\wedge\mathrm{demand}(c,d,\hat{\mathrm{P}})} \hat{\mathrm{P}}(c,d) + \mathrm{R}_{\mathrm{A}}(c,d).$$

The consumer c will follow a strategy where they demand d only if

$$\mathrm{demand}(c,d,\hat{\mathrm{P}}) \equiv \mathrm{R}_{\mathrm{I}}(c,d) \geq \hat{\mathrm{P}}(c,d) + \mathrm{C}_{\mathrm{A}}(c,d),$$

because consuming d will have a nonnegative impact on c's profit iff this condition is fulfilled. Similarly, the producer of d will follow a strategy where they supply d to a given consumer c only if the price satisfies

$$\mathrm{supply}(c,d,\hat{\mathrm{P}}) \equiv \hat{\mathrm{P}}(c,d) \geq -\mathrm{R}_{\mathrm{A}}(c,d).$$

A good d is produced if the producer manages to make a profit on a full cost basis, i.e.

$$\mathrm{produce}(d,\hat{\mathrm{P}}) \equiv \textsc{Tp}_{\mathrm{P}}(d) \geq 0.$$

Obviously, all choices of $\hat{\mathrm{P}}$ fulfilling all of these conditions for the same c and d result in the same allocations of resources other than money, and hence they realize the same social surplus.

This social surplus realized by a given set of prices $\hat{\mathrm{P}}$ can be quantified as $\sum_c \textsc{Tp}_{\mathrm{C}}(c) + \sum_d \textsc{Tp}_{\mathrm{P}}(d)$, which is

$$\begin{aligned} \textsc{RealSS} = \sum_{d|\mathrm{produce}(d,\hat{\mathrm{P}})} & -\textsc{Fc}_{\mathrm{I}}(d) \\ & + \sum_{c|\mathrm{supply}(c,d,\hat{\mathrm{P}})} -\mathrm{C}_{\mathrm{S}}(c,d) \\ & \qquad + \begin{cases} \mathrm{R}_{\mathrm{I}}(c,d) + \mathrm{R}_{\mathrm{A}}(c,d) - \mathrm{C}_{\mathrm{A}}(c,d) & \text{if demand}(c,d,\hat{\mathrm{P}}), \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

If we take the individual returns and costs faced by all participants together as a single "profit" maximizing economic entity, we can derive an upper bound on RealSS as

$$\textsc{MaxSS} = \sum_d \max\Big(0,\ -\textsc{Fc}_{\mathrm{I}}(d) + \sum_c \max\big(0,\ +\mathrm{R}_{\mathrm{I}}(c,d) + \mathrm{R}_{\mathrm{A}}(c,d) - \mathrm{C}_{\mathrm{A}}(c,d) - \mathrm{C}_{\mathrm{S}}(c,d)\big)\Big).$$

If we write this as a proportion,

$$\textsc{Pssr} = \frac{\textsc{RealSS}}{\textsc{MaxSS}},$$

then we can view this measure Pssr, the proportion of social surplus realized, as another evaluation measure for information retrieval.

Notes on Economics

This notion of an artificial market provides an approach to problems (x) and (xi) raised before, the problem of search costs, and the experience good problem. Note that, in the artificial market, each human consumer, has a retrieval system acting on their behalf as a fully automated agent. Thus, the search costs of the human consumer do not accumulate across all goods, but only across goods retrieved by the system, thereby reducing total search costs and providing a possible approach to problem (x). Furthermore, the fact that communication goods are experience goods is circumvented by having the computer agent "experience" the good on behalf of the human consumer in order to determine a reservation price, thereby providing a possible solution to problem (xi).

So far, we have considered only the resouce allocation problem without studying how money is allocated between producers and consumers within the space of possible efficient prices. For the purposes of a purely theoretic exploration of how information and attention can be allocated efficiently by the use of retrieval engines, seen as non-market pricing mechanisms, this question is perhaps not central. Money, in this case, could be a purely theoretic construct, rather than a real world financial instrument. For example it might be nothing more than a scale along which the costs and returns on information and attention are quantified for purposes of experimentation in information retrieval. However, given a suitable business model, nothing stops one from invoking real money with the pricing mechanism presented here. In this case, the allocation of money becomes economically a rather interesting one.

The range of efficient prices for a given good d and a given consumer c is, neglecting search costs and fixed costs, given as

$$\mathrm{R_I}(c,d) - \mathrm{C_A}(c,d) \geq \hat{\mathrm{P}}(c,d) \geq -\mathrm{R_A}(c,d).$$

This raises the question, where exactly in this range prices will be set according to the incentives of the retrieval engine, now seen as an economic agent itself.

There are three conceivable ways of assigning economic incentives to the retrieval engine: the retrieval engine could act on behalf of a producer, a consumer, or as an independent economic agent acting as a market maker.

If a retrieval engine is controlled by a monopolistic producer facing competitive consumers, such as the producers of information goods considered by Varian, prices would be set at the upper bound of the above condition. This would yield an efficient resource allocation, but the producer could extract all the surplus generated in the form of producer profit and drive down consumer surplus to zero.

If, on the other hand, a retrieval engine is controlled by a single consumer with great bargaining power over competing producers, prices would be set at the lower bound of the above condition. This would extract all the surplus in the form of consumer surplus and drive down producer profits to zero.

Finally the market maker could be a third party acting as an intermediary. If this market maker is a monopolist, while both producers and consumers are competitive, it could buy at prices at the lower bound, and sell at prices at the upper bound, thereby extracting all the surplus as profit, and driving down both producer profits and consumer surplus to zero.

At this point, it is important to note that the kind of competition that needs to be invoked for communication goods is not based on the traditional notion of perfect substitution. As pointed out by Varian, information is highly differentiated. A specific e-mail or a specific website will not have a perfect substitute. This is what we introduced as problem (ix) before. However, perfect substitutes are a rare phenomenon in any class of economic goods, so this does not come as a surprise, nor does it preclude the possibility of competition in general.

Our model already entails a notion of substitutability. If the consumer's behaviour is entirely determined by the profit function given before, the rate of substitution $\mathrm{RS}(c, d_1, d_2)$ at which a consumer c will substitute good d_1 for d_2 is simply

$$\mathrm{RS}(c, d_1, d_2) = \frac{-\mathrm{R}_{\mathrm{I}}(c, d_2) + \mathrm{C}_{\mathrm{A}}(c, d_2)}{+\mathrm{R}_{\mathrm{I}}(c, d_1) - \mathrm{C}_{\mathrm{A}}(c, d_1)}.$$

This means that, in a neoclassical-style interpretation, d_1 and d_2, even though they are not the same documents, could be considered perfect substitutes, if $\mathrm{RS}(c, d_1, d_2) = 1$ for all c. Otherwise an average rate of substitution $\mathbb{E}_c\{\mathrm{RS}(c, d_1, d_2)\}$ could be a useful analytic tool in determining a degree of substitutability between d_1 and d_2, and hence in analysing competition.

Notes on Retrieval

We will now show one example of how to instantiate this model for experiments on information retrieval. First of all we could make some simplifying assumptions to reflect the nature of a given retrieval application.

- We assume information is free. So we do not want our retrieval engine to take into account any incentives to show a document to anyone, just because someone has gone to the trouble of creating and indexing it. Hence, we set $\mathrm{FC}_{\mathrm{I}}(d) = 0$.
- We do not want our retrieval engine to take into account any incentives on the part of a producer of a given document to show the document to any particular consumer c. Hence we set $\mathrm{R}_{\mathrm{A}}(c, d) = 0$.

These two assumptions might be reasonable in a situation where no financial compensation is possible that would allow a producer to compensate a consumer for their attention. In this case the above assumptions would mean creating a retrieval strategy that acts on behalf of communication consumers and ignores the incentives of producers.

Furthermore,

- we approximate the search costs by a constant $\mathrm{C}_{\mathrm{S}}(c, d) = \mathrm{cs}$.

In general, these search costs would be determined by the amount of attention a user needs to expend in order to decide whether a given document is relevant or not. This does not seem unreasonable, when this decision is possible on the basis of a summary provided by the search engine. Examining such a summary would take a constant amount of time or attention.

Finally, let us make some more simplifying assumptions to make experimentation feasible. Say, for example, we were working in the experimental framework of TREC data (Voorhees and Harman, 2005). A TREC topic c reflects an information need. For each topic, we have a query q_c and a relevance judgement $\mathrm{rel}(c, d)$ for each document d. These relevance judgements are generally made with regard to the content of a given document and do not generally account for the amount of attention that has to be expended in order to read a document. For example, documents are not necessarily marked down just because they are long or difficult to read. Most retrieval engines evaluated in the context of TREC will have some natural notion of a retrieval status value to rank documents by. Thus, for any given query q_c and document d, we assume there is a retrieval status value $\mathrm{rsv}(q_c, d)$. – If no such notion of retrieval status value is available, an inverse rank or a percentile could perhaps be constructed from any given ranking. This experimental framework could then be translated into our evaluation model by making the following assumptions.

- Relevance judgements $\mathrm{rel}(c, d)$ reflect the return on information experienced by a consumer, i.e. $\mathrm{R_I}(c, d) = \mathrm{rel}(c, d)$.
- In the absence of any information on how much attention a given consumer c needs to expend on a document d, let us set $\mathrm{C_A}(c, d) = 0$.

At this point, our evaluation measure simplifies to

$$\mathrm{PSSR} = \frac{\sum_{c,d|\mathrm{rsv}(q_c,d)\geq 0} -\mathrm{cs} + \begin{cases} \mathrm{rel}(c,d) & \text{if } \mathrm{rel}(c,d) \geq \mathrm{rsv}(q_c,d), \\ 0 & \text{otherwise.} \end{cases}}{+\sum_{d,c} \max\big(\,0,\ \mathrm{rel}(c,d) - \mathrm{cs}\,\big).}$$

This seems intuitive, as $\mathrm{PSSR} = 1$ when $\mathrm{rsv}(q_c, d) = \mathrm{rel}(c, d) - \mathrm{cs}$ for all c and d. Depending on how exactly one translates relevance judgements to the cardinal scale used above for $\mathrm{rel}(c, d)$, and how one sets the value of cs on that scale, this measure reacts to different kinds of retrieval errors. Let us set $\mathrm{rel}(c, d) = 0$ on irrelevant documents, and $\mathrm{rel}(c, d) = +1$ on relevant documents and look at possible choices of cs.

If $\mathrm{cs} = 0$, the above measure is simply recall. Hence, the measure would react only to errors resulting in relevant documents not being retrieved. A retrieval engine that sets $\mathrm{rsv}(q_c, d) = 0$ on all documents, i.e. always retrieves everything, never makes an error of that kind. However, such a search engine would be less than useful. As the search costs approach zero, there is no point to using a search engine.

Next, consider the case where $\mathrm{cs} = 1 - \epsilon$ for an infinitesimally small $\epsilon > 0$. Now the numerator counts the number of irrelevant documents retrieved as a negative number. The measure would therefore react only to errors resulting in irrelevant documents being retrieved, which would traditionally be measured by precision. A retrieval engine that sets $\mathrm{rsv}(q_c, d) = -\epsilon$ on all documents, i.e. never retrieves anything, never makes an error of that kind. Again, the fact that such a retrieval engine would not be useful in practice is easily explained from the point of view of our economic model. As the search costs approach the value of the information contained in the relevant documents, there is no point to working with the document collection.

For any other choice of $\mathrm{cs} \in [0, 1)$, our measure would react to both kinds of errors, and the exact choice of cs can be read as a relative weighting of the two kinds of errors – similar in principle to β in the F_β measure (van Rijsbergen, 1979). However, the important advantage of our model is that, in response to problem (iii) raised earlier, we maintain an economic interpretation that can be used to set cs w.r.t. particular application needs.

For example, setting $\mathrm{cs} = 0.5$ means that the search costs associated with having to look at two documents are exactly the same magnitude as the returns a user gets from looking at one relevant document. Our measure now reacts to both kinds of errors.

- If $\mathrm{rsv}(q_c, d) = -1$ (i.e. the document is not retrieved), while $\mathrm{rel}(c, d) = +1$ (i.e. the document would have been relevant), then this document contributes a zero score to the numerator, and a $+0.5$ score to the denominator.
- If, on the other hand $\mathrm{rsv}(q_c, d) = 0$ (i.e. the document is retrieved), while $\mathrm{rel}(c, d) = 0$ (i.e. the document is not relevant), then this document contributes a -0.5 score to the numerator and a zero score to the denominator.

So it can be seen that these two cases decrease the value of this measure. Furthermore:

- If $\mathrm{rsv}(q_c, d) = 0$ while $\mathrm{rel}(c, d) = +1$, the numerator increases by $+0.5$, while the denominator also increases by $+0.5$. This would generally have a non-decreasing impact, where the measure remains unchanged if it already reflects a perfect score of 1, and where the measure increases otherwise.
- Finally, if $\mathrm{rsv}(q_c, d) = -1$ while $\mathrm{rel}(c, d) = 0$, then both numerator and denominator remain unchanged. So, in this case, the whole measure remains unchanged.

This behavior seems consistent with past experience on retrieval evaluation. The number of errors, which would decrease our measure, need to be compared to the number of relevant retrieved documents. When there are many relevant documents, in total, or when many documents need to be retrieved, in total, then we will naturally expect more errors. On the other hand, we do not want the measure to depend on generality, i.e. the documents that are irrelevant and not retrieved.

Also note that this formula implicitly microaverages over different topics, and that the zero-point of the scale for retrieval status values acts as a cost-based cutoff for ranked retrieval. There is no dependence on any rank-based cutoff chosen a priori. This offers a possible solution to problem (iv).

6 Concluding Remarks

In this paper, we framed the problem of searching for or filtering electronic documents as an economic resource allocation problem, while invoking retrieval engines as a non-market pricing mechanism. We have developed a mathematical model that has a dual interpretation in information retrieval and economics.

The economic side of this model helps us quantify the costs and returns involved in retrieving relevant or irrelevant documents and the costs associated with search itself. We believe that this model essentially presents one way of opening up the black box which is relevance, i.e. one way of modelling the interactions between information needs and collections of potentially relevant documents; between document utilities and rank-based or cost-based relevance thresholds; and the tradeoff between precision and recall. All of these aspects would enter traditional evaluation schemes in the form of parameters that have to be fixed a priori and that are often hard to interpret.

The information retrieval side of the model provides an economic mechanism for allocating attention and information where traditional markets would face problems arising from prohibitive search costs, the absence of competition based on perfect substitutes, and the experience good problem. Obviously, it is cheaper for a computer programme to examine a good than for a human. Therefore retrieval systems reduce search costs. The nature of competition for a communication good is that two goods are substitutes if they are relevant to the same information needs. In practice, this decision is often up to retrieval engines. Finally, to the extent that consumers trust the relevance assessments of search engines, search engines provide a way around the experience good problem by previewing and recommending goods to final consumers.

Ultimately, this paper is, to the best of our knowledge, the first attempt to model the economic efficiency of a given automatic retrieval engine. We believe this question should of utmost importance both to economists and researchers in information retrieval.

In section 2 we have listed many problems that arise in this context, and that have been pointed out in the literature before. At this point it is perhaps important to emphasize once again that these span broad fields in information retrieval and economics, and that we could not possibly hope to cover them in any great depth or to put forward a final solution to any one of them. Rather, the contribution of this paper, if any, lies in the theoretical groundwork developed, which touches on all of these issues. We hope that this helps to illuminate the interrelations between these issues, and that it might prove to be a useful point of departure for future interdisciplinary work at the interface of economics and information retrieval.

References

Baeza-Yates, R. A. and B. Ribeiro-Neto (1999). *Modern Information Retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Bakos, Y. and E. Brynjolfsson (1999). Bundling information goods: Pricing, profits, and efficiency. *Management Science 45*(12), 1613–1630.

Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems 30*(1–7), 107–117.

Buckley, C. and E. M. Voorhees (2005). Retrieval system evaluation. In E. M. Voorhees and D. K. Harman (Eds.), *TREC – Experiment and Evaluation in Information Retrieval*, Chapter 3, pp. 51–75. Cambridge, MA: MIT Press.

Cook, K. H. (1975). A threshold model of relevance decisions. *Information Processing and Management 11*, 125–135.

Cooper, W. S. (1973). On selecting a measure of retrieval effectiveness. In *Journal of the American Society for Information Science*, Volume 24, pp. pp 87–100 and 413–424. John Wiley & Sons.

Cooper, W. S. (1978, May). Foundations of probabilistic and utility-theoretic indexing. *Journal of the American Society for Information Science*, 107–119.

Cooper, W. S. and M. E. Maron (1978). Foundations of probabilistic and utility-theoretic indexing. *Journal of the Association for Computing Machinery 25*(3), 67–80.

Davenport, T. H. and J. C. Beck (2002). *The Attention Economy: Understanding the New Currency of Business*. Harvard Business School Press.

Fahlman, S. E. (2002). Selling interrupt rights: A way to control unwanted e-mail and telephone calls. *IBM Systems Journal 41*(4), 759–766.

Franck, G. (2007)). *Ökonomie der Aufmerksamkeit: Ein Entwurf*. Dtv.

Katter, R. V. (1968). The influence of scale form on relevance judgments. *Information Storage and Retrieval 4*, 1–11.

Lanham, R. A. (2006). *The Economics of Attention: Style and Substance in the Age of Information*. University of Chicago Press.

Lesk, M. E. (1969). Relevance assessments and retrieval system evaluation. *Information Storage and Retrieval 4*, 343–359.

Loder, T., M. Van Alstyne, and R. Wash (2004). Information asymmetry and thwarting spam. Available at SSRN: http://ssrn.com/abstract=488444.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy 78*(2), 311–29.

Page, L. (1998). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University.

Robertson, S. E. (1969). The parametric description of retrieval tests. *Journal of Documentation 25*, 1–27, 93–107.

Robertson, S. E. (1976). *A theoretical model of the retrieval characteristics of information retrieval systems*. Ph. D. thesis, University of London.

Robertson, S. E. (1977a). The probabilistic character of relevance. *Information Processing and Management 13*, 247–251.

Robertson, S. E. (1977b). The probability ranking principle in ir. *Journal of Documentation 33*, 294–304.

Robertson, S. E. (1997). The probability ranking principle in ir. In K. Spärck-Jones and P. Willett (Eds.), *Readings in Information Retrieval*, pp. 281–286. Morgan Kaufmann. reprinted.

Robertson, S. E. and H. Zaragoza (2007). On rank-based effectiveness measures and optimization. *Information Retrieval 10*.

Shapiro, C. and H. R. Varian (1998). *Information Rules: A Strategic Guide to the Network Economy*. Boston, MA, USA: Harvard Business School Press.

Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, Communication, and the Public Interest*. The Johns Hopkins Press.

van Rijsbergen, C. J. (1979). *Information Retrieval.* Butterworths.

Varian, H. R. (1995). Pricing information goods. In *Research Libraries Group Symposium on Scholarship in the New Information Environment*.

Varian, H. R. (1996). Differential pricing and efficiency. *First Monday 1*(2).

Varian, H. R. (1997). Versioning information goods. In *Digital Information and Intellectual Property*.

Varian, H. R. (2000). Markets for information goods. In *Monetary Policy in a World of Knowlege-Based Growth, Quality Change, and Uncertain Measurement*.

Varian, H. R. (2006). *Intermediate Microeconomics.* W. W. Norton & Company.

Voorhees, E. M. and D. K. Harman (Eds.) (2005). *TREC – Experiment and Evaluation in Information Retrieval.* Cambridge, MA: MIT Press.