Jon Dron, Chris Boyne, Richard Mitchell and Phil Siviter
University of Brighton, UK

Darwin among the Indices: A Report on COFIND, a Self-Organising Resource Base

Abstract: In this paper we report on the development and use of CoFIND (Collaborative Filter In N Dimensions), a web-based collaborative bookmark engine, designed as part of a self-organising learning environment to generate a list of useful and relevant learning resources. Users of CoFIND can add pointers to resources and rate them according to two types of category, 'topics' and 'qualities'. Along with the links and descriptions of the resources themselves, both topics and qualities are also entered by users, thus generating a resource-base and collective categorisation scheme based on the needs and wishes of its participants.

A topic is analogous to a traditional category whereby any object can be considered to be in the set or out of it. Examples of topics might include 'animals', 'computing', 'travel' and so on. Qualities, on the other hand are the things that users value in a resource, and most of them are (in English at any rate) adjectives or adjectival descriptive phrases. It is always possible to say of a quality that a given resource is more or less so. Examples of qualities might include 'good for beginners', 'amusing', 'colourful', 'turgid' and so on. It is the qualities that provide the nth dimension of CoFIND, allowing much subtler ratings than typical collaborative filtering systems, which tend to rate resources according to a simple good/bad or useful/useless scale. CoFIND thus dynamically accommodates changing needs in learners, essential because the essence of learning is change.

In use, the user enters a number of qualities and/or topics that interest them. Resources are returned in a list ordered according to the closeness of match to the required topics and qualities, weighted by the number of users who have categorised or rated a particular resource.

The more a topic or quality is used to categorise different resources, the more prominent its position in the list of selectable topics or categories. Not only do less popular qualities sink to the bottom of this list, they can also fall off it altogether, in a process analogous to a Darwinian concept of evolution, where species of quality or topic fight each other for votes and space on the list and topics and qualities are honed so that only the most useful survive. The system is designed to teeter on the 'edge of chaos', thus allowing clear species to develop without falling into chaotic disorder or stagnant order.

The paper reports on some ongoing experiments using the CoFIND system to support a number of learning environments within the University of Brighton. In particular, we report on a cut-down form used to help teach a course on Human-Computer Interaction, whereby students not only rate screen designs but collaboratively create the qualities used to rate those resources.

Mention is made of plans to use the system to establish metadata schema for courseware component design, a picture database and to help facilitate small group research.

The paper concludes by analysing early results, indicating that the approach provides a promising way to automatically elicit consensus on issues of categorisation and rating, allowing evolution instead of the 'experts' to decide classification criteria. However, several problems need to be overcome, including difficulties encouraging use of the system (especially when the resource base is not highly populated) and problems tuning the rate of evolution in order to maintain a balance between stability and disorder.

1 Introduction

In seeking a single resource in a world of exponentially multiplying resources, it is often helpful to know more about the qualities of that resource and how it has been valued by others. Traditionally, this is a role played by abstracts or individuals as critics and reviewers of papers for journals which form the various Seals Of Approval (SOAPs) we use to help us
discover resources we will appreciate. They are a means by which we discover, not only the kinds of resources we value, but also the value of those resources.

The World Wide Web has developed as a massive and unstructured body of useful, less useful, useless, and positively harmful information, a sea of data hiding nuggets of information embodying the knowledge of countless millions of publishers and participants in online communication. Search engines and web crawlers that sift through content, uncritically seeking keywords, may be very helpful in narrowing down the range of resources that we seek, but offer no clue as to the quality or reliability of those resources, and classification mistakes are the norm rather than an exception. Some improvement can be achieved by a process of self-classification using meta tags or standards such as PICS (Resnick 1997) or other XML schemas, but this too is potentially flawed. For example, a friend who produces sex education videos finds his work appearing next to every kind of pornography imaginable, even when sought with a discriminating set of search terms. Directories, as typified by Yahoo (Yahoo 2000), offer slightly better chances for my friend's work to be discovered, as they rely upon a human reviewer. Several problems arise, however, when we consider the relatively small number of resources that may be assessed this way, and the difficulties of relating one person's classification scheme to another's. Even where there is a similarity of outlook, our needs as searchers are unlikely to match those of the reviewer. Lakoff persuasively asserts that our entire world view can be characterised by the classification schemes that we employ (Lakoff, 1997), and that there are inevitably large and interesting differences between the classification schemes of different cultures, groups and individuals. Equally importantly, there are very few clues from the information that directories provide us with as to the qualities they possess that we will find valuable. At best, all that is available to us is a range of vague indicators such as frequency of links from other sites or names that we recognise and trust.

2 Collaborative Filters

Collaborative filters or recommender systems, such as GroupLens (Resnick, 1997), PHOAKS (PHOAKS, 2000) and Firefly (Chernenko, 1997) offer a solution to identifying valuable resources. Such systems assist the searching process by explicitly and/or implicitly gathering ratings of resources and combining them to provide recommendations. Automated Collaborative Filters, such as Firefly, go still further by identifying commonalities between users. The premise is deceptively simple: if you like 49 of the same books as I, it is probable that we will share a liking for the fiftieth. Similarly, if a hundred people mention a resource in newsgroups, it is likely that it will be interesting, perhaps even useful. Few of these systems broach the issue of categories, and whilst they are useful as means of discovering the value of relevant resources, they only scratch at the surface of the classification issues. An exception is RAAP, which is a collaborative bookmarking system for web sites which includes classification as one of the variables when matching users' needs (Delgado 1998). However, as with other recommender systems, every bookmarked resource is rated on a two dimensional scale, in this case based on popularity.

3 CoFIND

Like RAAP, CoFIND is a collaborative bookmarking system that incorporates a rating mechanism incorporating classification. Users are able to add categories (known in our system as topics) into which they may add URLs pointing to resources that they have found useful or valuable. As CoFIND is intended to be used in an educational environment, the ability to add topics allows students to collaboratively generate the structure of a course. Topics exhibit dynamic behaviour, rising and falling according to how much they are used. To identify the value of a resource, CoFIND uses discriminators in the form of metadata that we have characterised as qualities. Whereas existing collaborative filters such as Firefly and
PHOAKS rate resources on a single two-dimensional scale (good-bad, like-dislike, useful-useless, etc.), CoFIND encourages users to enter and rate according to a variety of qualities that are meaningful to them. We know that they are meaningful because they are added by the users themselves. It is these qualities that provide the \( n \) dimensions referred to in the name of the system.

Primarily designed for use in an educational environment (where needs are by definition likely to be novel and unlike previous needs), qualities are added to the system such as good for beginners, interesting, detailed, or simple. Once a quality has been added, existing and new resources may also be rated according to how well they match that quality. The value of a given resource is assessed according to the ratings received for the qualities that a user has selected as being of interest.

To prevent a proliferation of irrelevant qualities, the qualities are placed into an evolutionary landscape where they are in competition with each other. Only a single quality/topic combination may be selected at a time, and those that are not selected for some time fall to the bottom of the list and eventually 'die' of vote starvation, whilst those that are successful thrive at the top of the selectable list. The result is a list of qualities that have been found useful by users of the system within a given set of topics. Useful qualities provide useful lists of resources.

CoFIND is designed to allow a certain amount of 'stickiness' in qualities. Because highly rated qualities appear at the top of the list of available qualities, observation indicates that they are more likely to be used than those which require greater effort to select. For a quality to succeed it therefore has to be demonstrably more useful than the most successful. This is in keeping with the kind of evolutionary landscape found in natural systems, whereby success breeds more success, and allows qualities to better teeter on the 'edge of chaos'. Systems at the edge of chaos tend to display complex behaviour and self-organisation (Kauffman 1997). Significant work has gone into achieving this balance, with the current iteration of the algorithm based on a combination of frequency of use of a given quality to rate resources combined with frequency of selection of the quality when seeking resources. Were new qualities forced to compete on frequency of use, alone, then new qualities would never even reach the bottom of the list. Therefore, we artificially boost the initial position of a new quality on the list by giving it a novelty weighting. However, if it is not used much then the system self-organises and it very quickly sinks below those which are used more often. This rapid demise is caused by subtracting a number from the novelty rating every time another user logs into the system.

The algorithm used to display a list of resources is based on ratings given and the number of people rating a given resource, with the more popular appearing towards the head of the list, those rated more frequently having priority over those rated rarely. Thus if two people have given a high rating to a resource based on the currently selected quality it will appear before a resource rated equally highly by one person. However, if one person has given a high rating to a resource it will appear above resources given a lower average rating by many people. This gives newly entered resources a fighting chance in competition with resources that have been on the system for longer. However (in keeping with the general principles of self-organisation which underpin the system), the prominence of such resources means that they tend to attract more ratings and, should they be of less general use than the first rater believed, will quickly tumble to the bottom of the list.

CoFIND incorporates a threaded discussion mechanism that allows its users to add comments and discuss any resource. We provide this partially on pedagogic grounds but also as a means of providing more complex SOAPs than those provided by qualities alone. CoFIND is thus a self-organising classification system that develops organically out of the actions of the users of the system, not by an appointed classification expert or panel of experts. Depending on the users of a given CoFIND system, cultural and subject-based
differences and preferences become embodied or at least are brought into sharp relief, visible to all.

4 Experiments

Our main concern at the moment is to identify those qualities of resources that our students find valuable. The first trial discussed here was deliberately controlled so that we could concentrate on the development of qualities, without the complexity introduced by allowing the free addition of topics or resources.

4.1 The Discovery of Metadata Relating to HCI Issues

We performed this experiment largely to identify the issues involved and to explore the possibilities raised. We do not present this as an objective experiment but more as a piece of action research or proof of concept.

The week-long experiment involved a group of fifty University of Brighton level two students at the start of their studies of the Human-Computer Interface (HCI). The students were all very familiar with using the web but had not yet studied HCI in any detail. Twenty websites were carefully chosen by the tutors of the course to cover a wide range of topics and HCI issues. These resources were entered into CoFIND by the lead tutor. CoFIND had been customised so that students could not add new resources nor topics, but they were strongly encouraged to add qualities and rate the resources according to those qualities. Students were also free to use the discussion mechanism to comment on the resources if they wished.

The explicit learning objective was to encourage the students to reflect on what was good or bad about web pages, and, thus, by active experimentation to reflect on the issues surrounding HCI.

The system was seeded with four qualities (frustrating, hard-to-use, attractive, interesting content), mainly by way of example. These qualities were used by the lead tutor to rate the resources as he saw fit, so that from the start selecting any quality would produce a list of relevant resources. The experiment ran for a single week.

All students were asked to contribute to the system by evaluating at least ten sites and, if they felt the existing qualities to be inadequate to capture their thoughts, to add new ones. The students were not formally assessed on their contributions and it is interesting and a little sad to note that only seventeen students contributed over the required period. Over the course of a single week, these seventeen students added twenty qualities.

4.1.1 Some Preliminary Findings

The students generated the following qualities to describe the websites: messy, slow, boring, confusing, Navigation, Orientation, Appropriate Links, Innovative, Association Member, Author Info, fun, artistic, informative, indispensable, rich, self indulgent, Stupid, Unclear.

One hundred and twenty one votes were cast. By far, the most popular quality used to rate websites was attractive, which is not very surprising as it was amongst the first entered and has clear relevance to the subject matter. However, it is interesting to note that, following a minor cambrian explosion of qualities on the fourth day, that its massive popularity took a sudden nosedive at the expense of some other qualities, most especially artistic and interesting content. The evidence is inconclusive, but there is at least a suggestion that we were seeing the results of competition between qualities at work.

We were puzzled by moderate popularity of the quality, Author info. This is not a quality of websites that is often picked up as highly significant by writers on the subject. Although our results are far from unbiased, and are not experimentally controlled, this begins to hint that were we to extend the system to a wider audience that some surprising information
might be generated of use to web designers and those seeking to rate and classify web-based resources.

Three sites were outstanding in terms of numbers of ratings achieved: the University of Brighton's own site (understandable in the context); SNARG (a remarkable web art installation about which it is very hard to be neutral); and, to a lesser extent, the site for the Beano (a popular and long-lived British children's comic). The ratings achieved for each of these sites were very mixed, and help to show that a typical good-bad range of votes as used on conventional collaborative filtering sites fails to capture the subtlety of feelings that mark out a user's attitude to a site. SNARG, for instance, achieved very high ratings for positive quality of innovative, and the negative quality of confusing, whilst achieving low ratings for hard-to-use and frustrating (i.e., users felt that it was not hard-to-use and not frustrating). It would have been impossible to have gleaned this information using traditional two-dimensional numeric ratings and much more difficult to glean a consensus opinion using conventional SOAPs such as abstracts and critiques. We might have provided the students with pre-ordained qualities but we would never have come up with non-adjectival qualities such as Author info or Association member. In fact, it is fairly hard to see what was meant by these, but more than one student used both of these qualities to rate resources so it clearly had some meaning to them.

Several qualities (self-indulgent, stupid, and unclear) were entered but not used to rate resources. Whether this was due to the discovery that the qualities were not relevant or whether the students had problems using the interface remains to be discovered.

Only seventeen of the twenty resources were actually rated by the students. The three unrated resources were not hidden away at the bottom of the list of the returned resources for any of the originally entered qualities, so it is tempting to conclude that students were put off by the descriptions (free resource site, social work site and personal pages). Unfortunately, the mechanism for recording actual visits to sites suffered from a bug, so we are unable to check whether the sites were visited; but, there is a strong likelihood that they were not. As the task only required the students to visit and rate ten sites, and as their workload was high and motivation perhaps a little weak, it is likely that they failed to visit that many more than the required minimum. The implication of this seems to be that the first level of filtering takes place without regard to ratings, a matter of some concern to those seeking to achieve more effective website placement on search engines.

No one chose to make any comments or start a discussion, but as this was very much a task oriented exercise and there was no requirement to do so this is no surprise, if a little disappointing.

We have not found out what level of informal networking took place outside the context of the trial. As all the students were able to talk to each other throughout the week of the experiment it is likely that they discussed the sites and their reactions to them. We would not wish to discourage this as it is pedagogically very sound, but it almost certainly skewed the results.

4.2 Development of Topics with a Cohort of Masters Students

A full version of CoFIND has been available to a cohort of forty-five masters students studying computer networks over the past six months. It has not been well used for a number of reasons, most notably that this cohort has been something of a testbed for several less than wonderful iterations of the system. To avoid influencing their usage, they have intentionally been offered little help or incentive to use the system and were only given a few useful resources to get things started. Despite this, there has been a slow and unsteady increase in the topics, resources and qualities added to the system. We provided the students with a few resources to start with and the topics networks, and general computing. Topics added so far by the students are: health/safety info; travel; intranets; groupware; and internet. Students
ave added around thirty resources in those topics, which is not very encouraging given the length of time the system has been available, and it is interesting to note that on several occasions they have posted interesting resources to the cohort’s newsgroup in preference to CoFIND.

The small number of topics suggests, partly, that those that exist already were considered adequate, and certainly the majority of resources were added and rated within these topics. However, the main reason for the paucity of topics appears to have more to do with the cold-start phenomenon: it is inevitable that when a new topic is added that there will be few, if any resources categorised by that topic. This means that those seeking resources are unlikely to make much use of the topic as it will not return useful results. This causes a vicious circle where the popularity of the topic always remains low. We have grappled with this problem and not found a simple solution. RAAP gets around the difficulty by allowing users to import their Netscape bookmarks (including topic headings), and this is certainly an avenue worth pursuing, but it still requires a higher level of commitment on the part of the users than they may wish to give. We have also discussed positioning the system as a personal bookmarking system (accessible from anywhere in the world), but compared with the existing bookmark systems available within web browsers, the system is clunky, slow, and, above all, not available offline. We suggested this use to the students at the start of their course and they appeared enthusiastic, but no one appears to have used the system in this way, even though a personalisation option was added so that the main view they had was of resources that they had rated and topics and qualities which they had used in those ratings. A possible future route might be to kick start the system using AI techniques to extract category information about individual resources, but the amount of junk that such systems typically return may render the system as useless as it would be with no resources rated.

4.3 Forthcoming and Potential Uses of CoFIND

CoFIND is a general purpose tool that has many potential applications which we have only started to discover. Amongst these we have chosen to explore the following:

- Later this year we will be using CoFIND as a means of collaborative assessment, whereby qualities will be used as assessment criteria to grade work submitted by students as web pages;
- A group interested in developing IMS metadata for learning objects will use the system to collaboratively generate metadata to discover which categories are most appropriate for describing learning objects;
- CoFIND will be used a means of generating feedback on taught modules, including individual lectures and tutorials;
- An experiment that uses the qualities entered in CoFIND to categorise multimedia resources will be continued. This use of CoFIND helps to cope with the particularly intractable problems of finding non-text-based resources. Typical qualities that have evolved include sad, lots of red, funny, trees and black and white. This use is particularly interesting as it is inevitable that the border between topics and qualities becomes quite blurred.

5 Conclusion

CoFIND is an evolving system which, despite having been developed and used in a number of contexts for over a year, is still in an early stage of evolution. Like a word-processor, it has a wide range of applications, only a few of which have yet been explored. Many issues have yet to be resolved, particularly relating to encouraging use of the system. The cold-start phenomenon remains a serious problem. We have spent a lot of time playing with different algorithms to order topics, qualities, and resources, and have yet to get it exactly right. Adjusting the algorithm to dynamically cope with fluctuating usage where one week might see a flurry of ratings whilst another sees no activity at all, makes it difficult to
quickly adapt to changing needs. Making the changes too rapid results in a constantly changing and unstable system, whilst making them too static halts the development of the system. However, the results presented here demonstrate the potential for such a system to provide a means of filtering resources as well as to generate a consensus view of metadata.

References