Supporting Collective Intelligence on the Web: design, implementation and test of a self-organizing, collaborative knowledge system

Summary

The goal of this research proposal is to build a software system on the web, based on mechanisms of self-organization and adaptation, that efficiently helps its users to collaboratively develop a collective system of knowledge. The basic idea is to support the emergence of social or collective intelligence, i.e. develop an information system consisting of computers, software and people and connected by communication links, that is significantly more intelligent than its parts working in isolation. Rather than having each individual add his or her contribution to a knowledge store, the system should support the creation of a *shared mental map*, a network of interlinked concepts that constantly adapts to the needs of its users, both individually and collectively. This network would be similar to a collective neural network or "brain", that self-organizes and learns from from the way it is used.

The starting point is a group of people, who wish to tackle a particular issue. Such an issue may be complex and open-ended, but should be sufficiently focused to at least have agreement on what the most important topics are. Different members of the group can then propose different questions and answers concerning this issue. The software system will collect all contributions on a central server, make them available to the whole group, and create links between the contributions based on the explicit and implicit reactions of the users collectively. This will make the information easy to retrieve, highlight the best or most important contributions, point participants to the shortcomings of their contributions, and motivate them to add and improve on what they done. This constant reorganization and optimization of the resulting web will result from a model of social cognitive interaction, supported by a set of new algorithms for the interactive adaptation of websites.

Motivation

[this needs to be filled in further, but will be basically about why the web as it presently exists is not very reliable, why collective/social intelligence must be stimulated, which long term social and economic benefits we can expect from this approach, etc.]

Design and methods

We assume that the group of participants all have access via their browser to the World-Wide Web. Thus, they can contact the main server of the project, where their contributions are stored and processed. The set-up should both motivate the participants to address the outstanding issues in the best possible way, and facilitate their work by weaving together the different contributions so that the resulting knowledge system is as transparent and easy to use as possible.

Motivating the participants

We assume that the participants are interested in publicizing their own ideas, getting feedback from others, and getting answers to the questions they haven't as yet resolved. This should provide a basic motivation for collaboration. However, as everyone who has any experience working with a team of experts knows, this is far from sufficient to build up a comprehensive and coherent view of the problem domain. The complexity of the problem, its many different aspects or perspectives, the lack of time, the differences in opinion, and the lack of recognition or reward for one's contribution are just some of the many factors that make collaborative knowledge development difficult and inefficient. The set-up that we propose is intended to tackle these problems.

To focus attention and create a motivating structure, we start from a supply-demand model. First, we give every participant the opportunity to post those questions to the system that (s)he find interesting. Then, the system collects all question and asks the participants to evaluate each question (say, on a five point scale) indicating how relevant or important they consider each question. The questions can then be ordered according to their average score, so that the most important questions top the list. Using some of the algorithms which we will discuss further, the system can moreover use the evaluation pattern to determine which questions are similar, so that questions can be grouped or clustered into categories. This may suggest the creation of new questions, while a low score may indicate that some questions better be deleted. The process of refining the list of questions can thus go through a second or third round, to maximize clarity. The end result may be something like the "Frequently Asked Questions" (FAQ) lists that are one of the most popular formats to come out the Internet.

The list of questions together with their degree of importance represent the collective "demand" of the group, that is, the problems where the need for solutions is highest. The participants are then invited to answer whichever questions they feel competent to answer, however, preferably starting with those with the highest degree of importance. An answer takes the form of a short text, posted on the website, possibly with links to other documents on the same or a different website. We will call such an answer a "node" in the

conceptual network. A node will typically answer a single question, but may address other questions as well, as long as it remains short and focused. Nodes are linked into the question(s) they purport to answer. The whole of available answers determines the "supply" of solutions to the outstanding problems. The same question may be answered differently by different people, and it will be up to the group to decide which answers are most valuable.

We assume that the participants will prefer to have their contributions read, appreciated and reacted to. To achieve that, they should preferably write good answers to questions that are deemed important. Thus, "supply" will try to match "demand". In this stage of the process, the software system will support this intrinsic motivation by providing immediate feedback about the "popularity" of what one has written.

There are different methods to calculate the popularity of a web page. The simplest one consists merely in counting the number of visits to the page, but this is not very reliable as users cannot yet estimate the quality of a page when they decide to visit it. A better measure can be determined by explicitly asking visitors to evaluate the page after they read it, e.g. on a five-point scale. This, however, demands additional effort from the visitors, which they may not always be willing to spend. However, empirical results seem to indicate that there are almost as reliable evaluations implicit in the visitors' behavior, e.g. whether they print, save or bookmark the page. The simplest implicit evaluation is the time spent reading the page, and this has been shown to correlate well with explicit evaluations.

Explicit or implicit evaluations still have the drawback that not every visitor is similarly competent to judge about the quality of a page. The problem is that there is no objective way to determine who is most expert or authoritative on a particular subject. People become authorities or experts because they are considered as such by their peers. This implies a circular or *bootstrapping* mechanism: authorities get their status by being recognized by other authorities. Although this may seem paradoxical, recently a number of algorithms (PageRank and HITS) have been developed that solve exactly this kind of problem in the web: a webpage's authority is calculated recursively on the basis of the number and authority of other webpages that refer to it. Until now these methods have only been applied to the web at large, where the linking pattern is sparse and discrete, but mathematically they should work as well in a smaller website where the linking pattern is more fine-grained (as will be explained in the next section). Part of the project will be to test this out in practice, and observe in which ways authority measures are similar or different from the popularity measures sketched before.

In conclusion, frequency and duration of visits will provide a quick and flexible estimate of the quality of a page, while the linking pattern will establish a more long-term degree of authority. Contributors will be notified automatically about the score of their contributions on these measures,

especially after they have made changes to the page. This will stimulate them to constantly improve their work, and give them feedback about which changes produce the best result.

Underlying this dynamic is an implicit competition/cooperation model. Contributors try to improve their score, and to be better than others, and are thus competing. However, to improve their score they must rely on others, since their web page can only get more visitors through links with other webpages. Since links in the system will initially be bidirectional, linking to another relevant page will create incoming traffic and a flow of authority from this page. Thus, authors are motivated to seek out the most relevant links to other author's pages. In order not to link everything with everything else, however, the system must impose a maximum number of links per page (say, 10), so that authors will have to be discriminating in which pages they choose to link to. There is still the danger that they will all try to link to the same, most popular pages. This can be avoided by the algorithms for the automatic adaptation of links which will be discussed in a subsequent section. The fact that contributors can only improve their scores by relying on others makes the system inherently cooperative.

We expect that a "division of labor" will emerge in which authors describe the topic they are most expert in, but refer to others for the supporting concepts, that flesh out the details and context of their proposals. However, like in a market system there will in general still be different "suppliers" for the same product (answer to a particular question). These suppliers will compete not only for "customers" (readers of their contributions) but also for "subsuppliers" (other pages they link to). Although there will in general not be a single "winner" of the competition, there may be "losers", in the sense of contributions that get an unsatisfactory overall score. These may be retracted by their authors, "demoted" (moved down in the list), or even eliminated by the system. This creates a natural selection of contributions, so that only the best ones survive in the long term (although unsuccessful ones do not really need to be erased; they can simply be kept somewhere at the bottom of a long list, where they are unlikely ever to be seen by the casual user, but can still be retrieved by those who are really interested).

Self-organization of the network

The dynamic as sketched above is basically evolutionary, stimulating both variation (creation of new answers) and selection or "survival of the fittest" (selective presentation of the most popular or authoritative answers). However, the interconnection of the ideas is still left to the individual authors, who have only a limited understanding of the context in which their contribution would fit. With an ever growing collection of nodes, that are constantly improved by their authors, it becomes more and more difficult for individuals to decide about the most relevant links from their pages. A number of algorithms have

been developed by us that can tackle this problem, by letting the most appriate linking patterns emerge from the collective choices made by all users of the network.

The basic principle is an extension of the rule of Hebb for learning in neural networks: concepts that are used together or in quick succession become more strongly linked. We have demonstrated the usefulness of such an approach for reorganizing a website through a few, small scale experiments. The experiments used three learning rules: frequency, transitivity and symmetry. If a web user would move from page A to B and then to C, the frequency algorithm would increase the strength of the connections A -> B, and B -> C, the transitivity rule would increase A -> C, and symmetry would increase B -> A, and C-> B. We have shown that with these simple rules, a network of 150 nodes will self-organize quickly and efficiently from a random connection pattern to an associative network, where all related concepts are strongly linked to each other.

Recently, through theoretical reasoning, we have generalized this approach to a new, more sophisticated algorithm that in principle should reorganize a website even more quickly and efficiently. The main idea is that the time that a user spends reading a page (within certain limits, so as to avoid situations where a user opens a page and then forgets about it while answering a phone call) is a good measure of that user's interest for the page. Let us assume that a user spends quite some time reading page A, then browses quickly through pages B and C, then again reads D more attentively, and finally jumps quickly through E and F. In that case, we can assume that for someone who is interested in A, D is also quite relevant, but B, C, E, and F less so. In that case the algorithm will propose a strong connection from A to D, a weaker one from A to B and C, and an even weaker one from A to E and F.

The longer the delay between reading subsequent pages, the less we assume the pages to be mutually relevant, and therefore the weaker the strenght their connection get (strenght decays exponentially with the duration of the interval). However, the fact that two pages were read by the same user still implies that they have at least something in common and therefore the connection never becomes completely zero. This is basically an extension of the transitivity rule, which now is no longer limited to pages that are at most two steps away, but applies with decreasing force to pages that are an unlimited number of steps away from the starting point. The equivalent of the symmetry rule is simply to add a constant fraction of the connections strengths (e.g. A -> D) generated by this rule to the inverse connection D -> A.

The algorithm can be made more reliable by replacing the truncated durations of visits by explicit evaluations of the relevance of a page. However, as we noted earlier, requiring the user to explicitly evaluate each page demands an effort that few will want to spend when merely browsing through

the web. Therefore, explicit evaluations will merely function to complement and check the reliability of the evaluations implicit in reading duration.

Although we are convinced that this algorithm will work at least as well as the earlier one that was shown to be effective on a small scale, the new algorithm too will have to be tested, preferably on a realistically large and complex website. This will also help us to determine the best values for the different parameters (e.g. speed of exponential decay, stable contribution of user, contribution of inverse connection) included in the formula for calculating the link strength.

The result of applying this algorithm to the web will be the generation of a square matrix whose elements are the strengths of the connections between any two nodes (corresponding to the rows and columns of the matrix) in the network. Every additional visit to the network will slightly change this matrix, so that it constantly adapts to the changing use patterns. The applications of this matrix are manifold:

1) from the matrix a list of recommended links can be derived for any given page; these are the links with the highest strength, in the order of their strength. Thus, the matrix determines which page is linked to which other pages.

2) by considering the mutual connection strength of links between nodes as a measure of similarity (or inversely, of distance), nodes can be clustered into groups of related topics. This can produce an automatic classification of subjects in the web, and generate a set of indexes for particular subdomains. Thus, new overall subjects or domains will spontaneously emerge from associations between different contributions.

3) by representing the interest profile of a particular user as a vector of activations (the activation of any node in the list is proportional to the degree of interest that the user has shown), individual recommendations can be generated for every user through "spreading activation": the activation vector is multiplied repeatedly with the connection matrix until the resulting product vector has stabilized.

Personalized recommendations

The latter method is a very powerful way to help users find the pages that are most interesting to them, even if they don't know what is available in the web, or cannot explicitly formulate their interests. It suffices that the user would browse the web for the system to collect data on which pages the user found interesting (the longer the reading time, the more interesting). This determines an activation vector which is used to produce a recommendation for those pages the user hasn't visited yet, but which, on the basis of the recorded paths of other users, seem most strongly related to the whole of the pages that the user liked. Since the activation vector changes with every page that the user visits, the recommendations constantly adapt themselves to ever better reflect

the user's present interest (an exponential decay factor can be used to reduce the activation of pages visited a while ago).

Again, the recommendations can be made more reliable by allowing the user to explicitly evaluate the components of his or her interest profile: the system can simply show a list of the pages visited, each with a graphical depiction of their estimated degree of relevance. The user can then edit the relevance score, e.g. by clicking one out of a series of radio buttons that represent increasing degrees of interest.

Such a recommendation system can be viewed as a search engine without words, since it will retrieve pages the user is looking for without the user specifying any keywords. The method would therefore work equally well with webpages containing purely non-verbal information, such as photos or sounds. Still, association-based recommendation can be used in conjunction with keyword searches, e.g. by adapting the order of the search results to better reflect the user's interest. Spreading activation is a general model of associative thinking as it happens in the brain. As such it is much more sensitive and flexible than formal deduction or exact retrieval of words or patterns. The system can also be programmed to automatically warn users of any changes in pages that the user found interesting.

These individual recommendations can be seen as special cases of the overall ordering of nodes by the system. The degree of activation achieved by spreading activation from a homogeneous vector (where all nodes are initially considered to be similarly interesting) determines a kind of unbiased recommendation, for the novice user who hasn't visited any pages yet. The activation of nodes in this unbiased list plays a role equivalent to the "authority" of nodes as calculated by the algorithms mentioned earlier. Different approaches for calculating these overall activations and different values for the parameters in these calculations will have to be tested out in order to find the best model.

Interaction between the participants

If we wish the system to improve not only its structure but also the content of its pages, we should stimulate participants to criticize and comment upon other partipants' contributions. Therefore, the system should provide a full capability for *annotation*: participants should be able to attach comments to any page or section of a page in the system. These comments can themselves be equivalent to "nodes" or pages in the web, that simply link to the page they refer to. However, comments do not need to contain any contribution that can stand on its own: they might consist simply of a one line statement saying "this is wrong".

It should also be easy for other participants to expand on, or reply to, such annotations, so that a "thread" of discussion can evolve within the system, attached to a particular node that determines its subject. Such threads can be

similar to a newsgroup or mailing list discussion, although they should be restricted, so that they remain on topic, instead off wandering off in any direction that seems interesting. Of course, starting a new topic must not be discouraged. The system should simply make a clear distinction between new contributions, which require a page of their own, possibly linked into the discussion page that triggered it, and on-going arguments and counter-arguments that all refer to the same basic question, and are therefore sequentially added to a single web page.

If we want to stimulate discussion as well as the introduction of new, unfinished, or controversial ideas, it seems useful to allow participants to remain anonymous. Thus, contributors will not be afraid to express ideas that may get a lot of criticism. Since the whole mechanism is based on natural selection, there is no reason to *a priori* select contributions based on quality requirements: poor ideas will automatically be pushed to the bottom of the heap. Therefore, it is best to have a maximal variety, and this may include wild, as yet poorly formulated, or dubious ideas. If there is any value in the idea, this will hopefully be clarified through the subsequent discussion, and the selection mechanism will make sure that the authority of the page increases as the remaining problems are eliminated. To reduce participants' fear that their half-baked ideas would badly affect their reputation, the system should always allow them the option of remaining anonymous. This option should probably be the default for criticisms of other people's ideas.

Better than pure anonymity are pseudonyms, where the system knows who wrote the piece, but no one else can find out. If an initially risky idea then afterwards develops into one that is well-appreciated, the author will have the option to drop anonymity, and the system will simply replace the pseudonym by the real name. Thus, the authors will receive due recognition for their good ideas, while being shielded from personal criticism for the bad ones. Since the system knows which author hides behind the different pseudonyms used, this form of anonymity does not prevent detailed personal feedback: the system can notify authors of reactions to, or evaluations of, their contributions independently of the name under which the contribution was entered into the system.

Testing and evaluation of the system

Once the main parts of the system have been implemented, and users have been sufficiently active for the system to have learned from their activities, the different components of the system must be evaluated to see in how far they achieve the intended aims, and how they could be improved. This evaluation and testing can go on in parallel with the continuing use of the system, since the data produced by one component can be compared with data produced by another, independent component, without need to halt the on-going

developments. This will allow us to both get fine-grained, quantitative evaluations of specific components of the system, and get a more global, qualitative apprehension of its overall development.

Evaluating the demand and supply model

The effect of increased demand (higher score for particular questions) can be evaluated by calculating the correlation between the score of a question and the number of answers it gets, and the delay between the addition of the question and the creation of the first answer to it. Another method of evaluation consists in polling contributors informally about how they have been influenced by the score of the questions and the feedback they got on their answers.

Comparing the different ranking methods

We have proposed different methods to calculate the "importance" or "popularity" of a page, such as number of visits, duration of visits, and "authority" calculated recursively from links. The results of these scores for the pages can be mutually correlated. If the correlation between two scores is poor, this means that they are measuring different aspects, and then a more detailed analysis may be needed to find out where and why they differ. All computed scores will also be correlated with an explicit evaluation by the users, which seems generally more reliable (although it could be argued that in some cases the computed scores (e.g. authority) would uncover additional properties). This will allow us to establish how reliable each method of scoring is, and how the different methods could be combined to produce and even more reliable score.

Evaluating the quality of learned links

The continuously adapting matrix of cross-connection strengths between pages determines a list of most recommended links for any given page. The quality of these recommendations can be tested simply by asking people to compare the relevancy of links that are, unknown to the user, created in different ways: system-generated, provided by the author, or randomly chosen. If our assumptions are correct, the system-generated links should on average score significantly better than the random links, and at least as well (and probably better) than the author-provided links. The exact score can be used to fine-tune the algorithm (e.g. by adjusting the value of the different parameters).

Evaluating the quality of clusters

A similar technique can be used to evaluate system-generated clusters of related pages: users are asked to evaluate the quality of various collections of links, some system-generated, some author-generated, some random. This

again will provide us with feedback that will allow us to improve the clustering algorithms.

Evaluating personal recommendations

Again, a similar method can be used to test the quality of the personal recommendations the system generates on the basis of a user's interest profile. Now and then, the user is asked to explicitly evaluate the quality of some of the recommendations that (s)he received. Unknown to the user, some of these recommendations are based on the personal profile, some are random, and some are merely based on overall popularity. Our hypothesis is that personalized recommendations will score much better than random ones, and somewhat better than those based on overall popularity. The differences in score can again be used to improve and fine-tune the spreading activation algorithm and the personal profiling method which generate the recommendations.