Web Intelligence and Agent Systems: An International Journal 6 (2008) 441–452
DOI 10.3233/WIA-2008-0150
IOS Press

Web search engine working as a bee hive

Pavol Navrat *, Martin Kovacik, Anna Bou Ezzeddine and Viera Rozinajova

Slovak University of Technology, Faculty of Informatics and Information Technologies, Institute of Informatics and Software Engineering, Ilkovicova 3, SK-842 16 Bratislava, Slovakia

Abstract. A new approach to web search that is based on a bee hive metaphor is presented. We proposed a modified model of a bee hive. Our model comprises of a dance floor, an auditorium, and a dispatch room. We have shown that the model is a true model of a bee hive in the sense it simulates several kinds of its typical behaviour. However, more importantly it is a simple model that describes some processes taking place in web search. Our model incorporates also several kinds of uncertainty. Experiments show that uncertainty increases robustness of the search and makes it actually more efficient.

Keywords: Web search, bee hive, Page Rank, recommender system

1. Introduction

The world wide web contains a lot of information, some of it undeniably useful. Although it is definitely not the case, as the popular saying goes, that "everything is on the web", the volume of information that is somewhere on the web is growing continuously, opening, by the way, both opportunities and threads (for a discussion on some of the issues in the e-learning application of the web, see e.g., (Chuda, 2007)). The fact that the information is somewhere on the web does not imply automatically that it is available (i.e., it can be accessed, retrieved or found by search) immediately, or that it will be available with the given methods at all. Role of search, viewed as a more general process, is crucial.

From this more general perspective, there should be left no directions unexplored, if they could only contribute to some improvement of the web searching process (current research explores e.g. support to the information retrieval process (Koval, 2002), semantic web (Hluchy, 2007; Matusikova, 2007; and many others), as well as various other approaches e.g. (Ye, 2005; Lei, 2005; Zhang, 2004)).

On the other hand, it may be useful to narrow the perspective so that a productive procedure emerges. For example, it may be useful to consider recommending information found on the web, or even more specially recommending a case from a base of cases.

We propose to explore possibilities open by considering behaviour of social insects. Social insects (ants, honey bees, termites, wasps etc.) show a remarkable level of social behaviour. In particular, they communicate, albeit in a very elementary way, with each other. For example, they may communicate regarding the location of food sources. They even collaborate towards achieving some goal. In particular, they may collaborate regarding bringing the food back once it is found. They distinguish themselves by their organisation skill without any centralised control (Gordon, 1996). The interactions among individuals, between the individuals and the environment along with the behaviours of the individuals themselves allow organisation to emerge.

Our hypothesis is that the behaviour the particular social insects show is an instructive inspiration to develop a model of searching the web for a source of information.

The rest of the paper is organised as follows. In Section 2, the notion of a bee hive is briefly introduced. Bee hive model and its usage in recommender systems is discussed in Section 3 and 4. Proposed modification of this model is presented in Section 5. The next section is devoted to modelling information retrieval from web including possible contribution to information recommending or filtering. Sections 7 and 8 deal

*Corresponding author. E-mail: navrat@fiit.stuba.sk.

with evaluation of web page relevance and possible benefit which could be brought by our bee hive model. Conclusion and future work is the subject of Section 9.

2. A bee hive

Behaviour of honey bees, *Apis mellifera* has been studied extensively by natural scientists, e.g. sociobiologists and behavioral physiologists. One of their aims has been, from our perspective, to formulate a model of bees' behaviour, specifically of their collective nectar source selection. Basically, there are various approaches possible: e.g., to build a simulation model, or an analytical model expressed using differential equations. A mathematical model of (Camazine, Sneyd, 1991) has proven to be especially inspiring. A very elaborated model has later been presented by (de Vries, Biesmeijer, 1998).

Honey bees are social insects who live in colonies. Bees collect food (nectar) from sources distanced up to 10 km from the hive. The colony uses simple rules to dispatch bees towards the best nectar source available. In this process of foraging, the bees return with nectar and the information about its source. Since the sources are not constant, but new ones appear here and there from time to time, as well as the existing ones become exhausted, the information about nectar sources is crucial. The colony keeps on adapting itself to the ever-changing situation. To be able to do so, and to maintain a sufficient influx of food, some division of labour force between exploring the surrounding countryside for new sources and exploiting the existing ones is necessary. In other words, there are bees, which go on a foraging mission, but there may be other ones, which are explorers at the given moment.

The foraging bees are able to pass the information they have on the location of the food source they visited onto other bees, and other bees are able to receive that information. Important role in the process of communication plays a waggle dance. Many have studied the mechanism of the waggle dance, the information that is communicated and the way it is actually done (von Frisch, 1967). It is hypothesized that distance and direction are encoded in the waggle dance, although they may be not the only types of information that the foraging bees are guided by. Discussions about these topics, no matter how interesting they may be, are clearly outside the scope of this research. We shall assume that the waggle dance is a means of communicating a food source. The dancer lets know not only distance and direction of the food source; duration of dancing is influenced by the quality of the source.

3. Bee hive modelling

There have been efforts to develop a model of a bee hive. The theoretical model by (Camazine, Sneyd, 1991) supposes that bees either unload nectar they have just brought from source s, or dance promoting source s, or continue foraging at source s, or observe dancers promoting sources $s_1, s_2, \ldots$, or follow a new source from $s_1, s_2, \ldots$, after observing. Each of the choices can be characterised quantitatively by probabilities.

The authors experimented with the model. Their experiment aimed at investigating how colonies choose among nectar sources (Seeley, 1991). In the vicinity of a colony, there were placed two nectar sources, one of them 400 m to the north, the other one 400 m to the south. 12 bees were trained to fly to the north source, 15 other bees to the south source. The sources were of different quality. The south one was better (sugar concentration of 2.5 units) than the north one (1.0 units) initially. At noon, however, the sources were swapped. Next morning, the sources were swapped again, and at noon once again. Empirical observation showed that number of bees foraging for the better source was increasing in time, whereas number of bees foraging for the worse source remained low. After the quality reversal, the trends reversed, too. Simulation experiments with the model yielded a degree of similarity with the empirical observations (Fig. 1).

Some years later, de Vries and Biesmeijer (1998) attempted to build a different simulation model that would as a minimum give experimental results similar

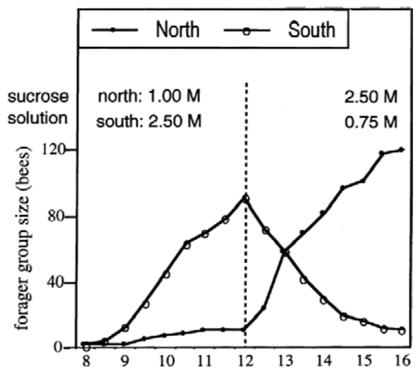

Fig. 1. Result of the experiment aimed at investigating how colonies choose among nectar sources (Seeley, 1991).

to the original empirical experience of (Seeley, 1991). However, their ambitions went beyond that. The simulation model allows studying rules, according to which bees behave. Indeed, they succeeded to formulate a set of rules. Their concept of forager bee is more refined. A potential forager bee is initially "unemployed". She does not know about nectar sources outside. Sooner or later, she will fly out for a nectar source. She can fly out "spontaneously" as an explorer (scout) i.e. not having a definite destination. Or she can visit dance floor, observe waggle dance performers and become a recruit. Then, she starts foraging having stored approximate positional information in her memory, whereas the scout starts without it.

Their model has been successful in faithfully simulating some kinds of collective foraging behaviours. Their range is largely determined by the set of rules. If the authors extended it, the range might become larger. In order to model bee hive foraging patterns of behaviour more comprehensively, several other extensions could be proposed.

We have developed another model that allows to simulate some foraging behaviours of a bee hive. However, since our aim has been to create a model that would bring some new ideas to methods of searching the web in a general sense, we shall introduce it in the next section.

4. Bee hive metaphor

Honey bees behave in a very interesting way achieving remarkable results, viewing them from the problem solving perspective. To consider bees in this context is a relatively new idea (Tovey, 2004; Lorenzi, 2005).

One of the earliest mentions actually dates back to 1986, when Bullock, Dey and Reilly (1986) suggested in a short note a bee hive model for heterogeneous knowledge in expert system, replacing blackboard by a more de-centralised scratchpad. Dornhaus (1999) studied methods of multi-agent modeling using as an example a bee recruitment. In a conference poster by Schultze (2002), bee foraging has been applied to enhance the collaborative process of filtering information.

One of the most recent projects by Lorenzi et al. (2005) makes use of the bee hive metaphor in a case-based recommender system. Their idea is to use bee dance to retrieve the most similar case to the user's query. They adopted model of Camazine and Sneyd (1991) and combined it with case-base reasoning. Case corresponds to a nectar source. Case base is a set of nectar sources. In a standard case-based recommender, a user query is compared with all cases in the case base. The most similar is returned to the user. In their recommender, bees decide on abandoning or continuing to visit the case depending how similar is the query to the visited case. In other words, the probability of abandoning the source decreases with greater similarity between the query and the case. Bees start foraging randomly and after some time, a case with the clearly highest number of bees emerges.

5. Modified bee hive model

Let us now proceed to presenting our model of a colony of bees living in a hive. Its rationale is the following. We want to explore possibilities of defining the process of searching, retrieving etc. information in a way that is similar to the process of searching, bringing etc. food to the bee hive. Intentionally, this is a rather broad formulation, since we want to research in various directions, not making assumptions from the start what will be the best one, if any. First thing to do was to devise a model of a bee hive, that would both be consistent with the biological original and suitable for experimenting in information retrieval.

Therefore, there are two aims regarding our model: it should model behaviour of biological bee hive and it should model information retrieval. However, our main objective is a possible contribution to information retrieval. We shall devise a bee hive model, but we do not obviously aim at a contribution to research in biology. It will suffice if our model exhibits a degree of similarity with the bees' foraging behaviour. One of the properties of our model should be simplicity.

We took an initial inspiration from the model of Lorenzi et al. (2005) who in turn were inspired by Camazine and Sneyd (1991). Our model of the bee hive comprises of a dance floor, auditorium, and dispatch room. We enriched their model by introducing a dispatch room (Fig. 2).

The dispatch room brings additional flexibility to the model. It is an explicit reflection of the hypothesis that some bees in the auditorium may not be ready to follow any one of the advertised sources. They may either not become enthusiastic for any dancer within some reasonable time, or there may be not any dancers to observe within some reasonable time. Therefore, they simply decide better to go foraging blindly than run into a danger of starving in the auditorium. More-

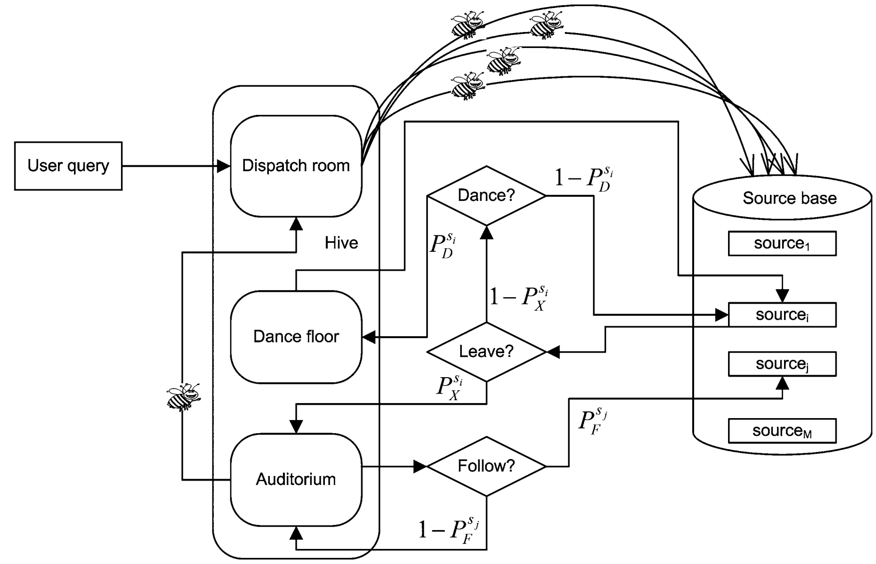

Fig. 2. The bee hive comprises of the *dance floor*, *auditorium* and *dispatch room*.

over, bees take off from the dispatch room initially, when a new query arrives.

The model is flexible due to several its attributes which are parametrised. Parameters of our model are the following:

overall number of bees N (*BIOR* + *BISB*) in the initial state, it consists of bees in the auditorium (*BIOR*, bees in the observing room), and those which start foraging (*BISB*, bees in the source base),

initial distribution of bees between observers (*BIOR*) and foragers (*BISB*),

maximal time allotted for dancing for a source (maximal dancing time *MDT*),

maximal time a bee can spend in the auditorium (observing time *OT*).

It should be noted that the time a particular bee dances depends on a quality of the source (it equals *MDT*$*$*quality*). The observing time, on the other hand, is either the maximal one or is shorter if some dancer before the maximal observing time elapsed attracted the particular bee.

The process of searching for the best source starts when a user query arrives. In response to that, *BISB* bees take off from the dispatch room randomly, i.e. each of them is sent to a randomly chosen source. Each bee, having visited the source, evaluates it with respect to the user query. The quality Q_i of a source s_i is defined as a measure how much the source matches the query.

Having returned to the hive, the bee decides if she abandons the source she just visited (depicted as the selection diamond *Leave?*). The model defines a probability $P_X^{s_i}$ ($P_X^{s_i} = 1 - Q_i$) that the bee abandons the source s_i. The probability of not abandoning the source s_i is obviously $1 - P_X^{s_i}$. If the bee decides to abandon the source, she immediately enters the auditorium.

If the bee stays with the source, she has again two options. She can either try to attract some fellow bees for the source s_i or she can take off for another visit to the source s_i (depicted as the selection diamond *Dance?*). The model defines a probability $P_D^{s_i}$ ($P_D^{s_i} = 1 - P_X^{s_i}$) that the bee goes dancing for the source s_i. The probability of not going to dance is obviously $1 - P_D^{s_i}$. In that case she flies to the source s_i. If the bee decides to dance for a source, she immediately enters the dance floor. The better the source quality, the longer she dances for her source. However, the dancing time cannot be longer than *MDT*. After the bee completes her dancing engagement, she flies to the source which she was trying to advertise by dancing.

Bees in auditorium observe dancing bees. They try to make a decision which dancer to follow for a source. An observing bee chooses a source by choosing randomly some dancer. Let us assume this dancer advertises a source s_j ($j \in \{1, \dots, M\}$). She will fly to this source with a probability $P_F^{s_j}$. The model defines the probability $P_F^{s_j}$ as the number of bees dancing for the source s_j divided by the number of all dancers. The probability that the bee will not fly to the source s_j is $1 - P_F^{s_j}$. In that case, the bee will choose randomly another dancer.

If the bee does not choose a source within the alloted time, she moves from the auditorium to the dispatch room. From it, she flies to a randomly assigned source. The whole process repeats as described here.

The model just described here is a simplification of models studied by biologists. Still, we hypothesize that it is, at least in some respect, a model of a bee hive. So we do not attempt to model all the aspects of a bee hive's behaviour, having in mind our main goal of developing a model of web search. We tried to perform a similar experiment as the one published by (Seeley, 1991), because this experiment has been used in the literature as a kind of testbed for bee hive models.

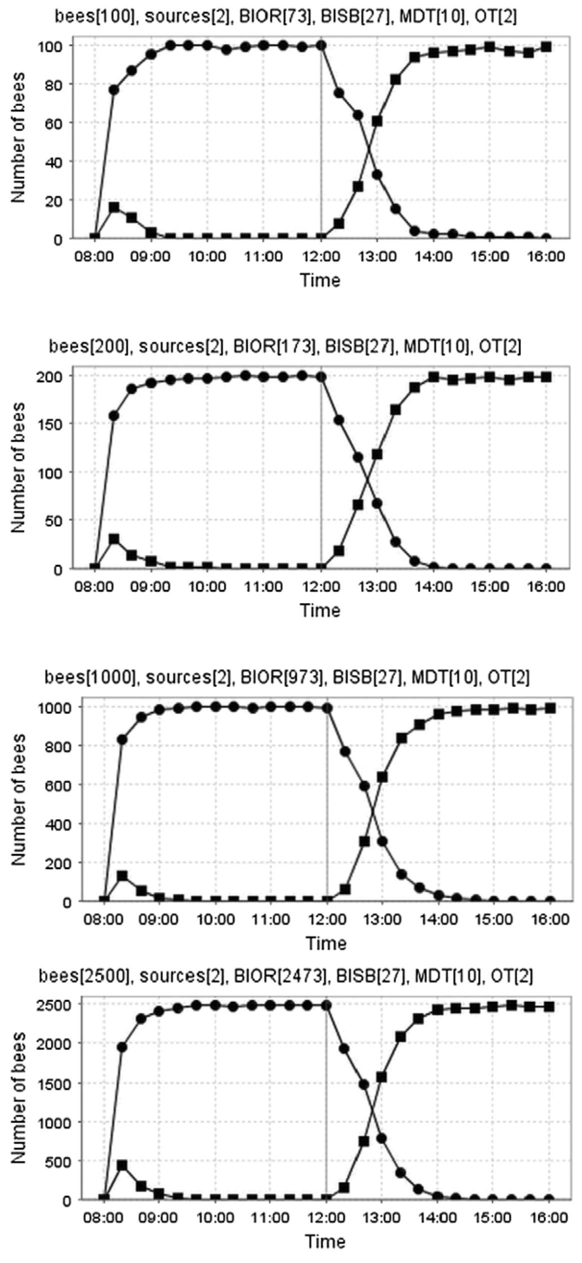

Fig. 3. From 8 am, bees forage for two sources. The one in the *north* (depicted by a rectangle) is initially worse, the one in the *south* (depicted by a circle) is better. At high noon, the sources are swapped. The bees react accordingly.

In this experiment, our model does not describe uncertainty that experiences a bee when she flies towards a source. For the time being, let us accept the results that show the principal similarity with the experiment, abstracting from the steepness of the curves etc.

Some bees have been "trained" initially. They knew, when taking off from the dispatch room, where to fly: 12 bees were trained to fly to the north source, 15 other bees to the south source, totalling 27 bees in the source base. In our experiment, the bees were able to attract other bees to join then foraging for both sources. Soon, however, the better source prevailed. After the sources were swapped at high noon, the bees were able to redirect their mission targets. This is in full conformance with the original experiment.

We present results of our experiments with different numbers of bees in the hive. The basic trend is in all experiments the same. Higher number of bees makes the behaviour smoother.

6. Modelling information retrieval

We hypothesise that the bee hive metaphor could prove useful in devising methods of web search, and in particular, of information retrieval from web. We investigate several possible directions of research, including a possible contribution to information recommendation or filtering. At this stage, we wish to present results that may shed some additional light into these considerations, stimulating at the same time further discussion.

Documents on the web have usually several attributes. Some of them can have only one out of two values, some can have value from an interval of real numbers. This is a fairly general assumption, although we definitely do not claim it covers all possible cases. Let us assume, for the simplicity sake, that the binary valued attributes can be either 0 or 1 and the value 1 is the one matching the user query. Let us further assume that the real valued attributes can have value from the interval [0,1] and higher value better matches the user query.

We have performed a series of experiments, out of which we shall present here only two. In both experiments, we assumed that the sources have three attributes. Two of them are binary and one is real. In the first experiment, a bee was able to evaluate the source how close is it to the user query by considering all the three attributes. In the second experiment, we tried to create a situation when some bees will during foraging use only the first attribute, some will use only the second attribute and some will use only the third attribute.

Results of the first experiment are shown in Fig. 4. Bees find the best source and massively forage for it.

The sources are (listed in the order of quality): source12[0.85], source39[0.65], source54[0.56], source98[0.54], source43[0.36], source11[0.33], source63[0.33], source87[0.33], source18[0.32], source24[0.32], source92[0.31].

As we can see, bees are more confident in recommending the best source if there are more of them foraging. But even a bee hive with only 10 bees was able to recommend the best source out of 100 candidate sources.

Results of the experiment shown in Fig. 4 are hardly surprising. Indeed, once we have established a basic similarity to the bee hive, as demonstrated in Fig. 3, it should have no difficulty in demonstrating a strongly decisive recommendation of the bees in this experiment, where the only additional difficulty was a slightly more complicated evaluation of the similarity between the query and the source. Still, please note that in the beginning, bees go for various sources and only later unite on the best one. In other words, the bees truly search the space of sources (and are quite good at it). We include this experiment also to have its results available for comparison to the next one.

In the second experiment, the sources are (listed in the order of quality):

source12[0.85], source39[0.65], source54[0.56], source98[0.54], source43[0.36], source11[0.33], source63[0.33], source87[0.33], source18[0.32], source24[0.32], source64[0.32], source92[0.31], source78[0.26], source89[0.23], source7[0.20], source9[0.17], source95[0.11], source40[0.09], source91[0.08], source35[0.07] etc.

In the second experiment, each bee is capable of evaluating one single attribute. Each source has three attributes so any evaluation made by an individual bee is only a partial one. Some bees contribute to recommendations by evaluating only the first attribute, some only the second attribute and some only the third attribute. A source with better evaluations in each attribute will be rewarded by higher number of combined bee visits. Objective of our experiment was to see if this hypothesis deserves further attention or it should be rejected.

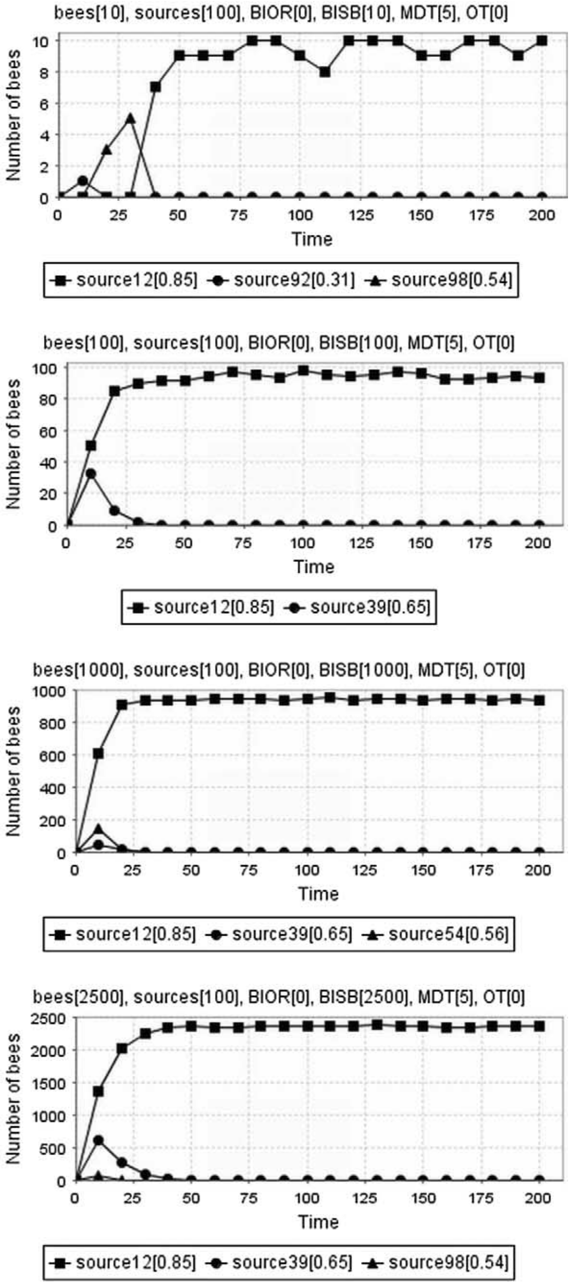

Fig. 4. Various numbers of bees forage for 100 sources. Sources have 3 attributes, bees evaluate all of them.

Results of the experiment are in Fig. 5. Bees were able to recommend the best source in two cases (100, 1000 bees). They nearly missed the best recommenda-

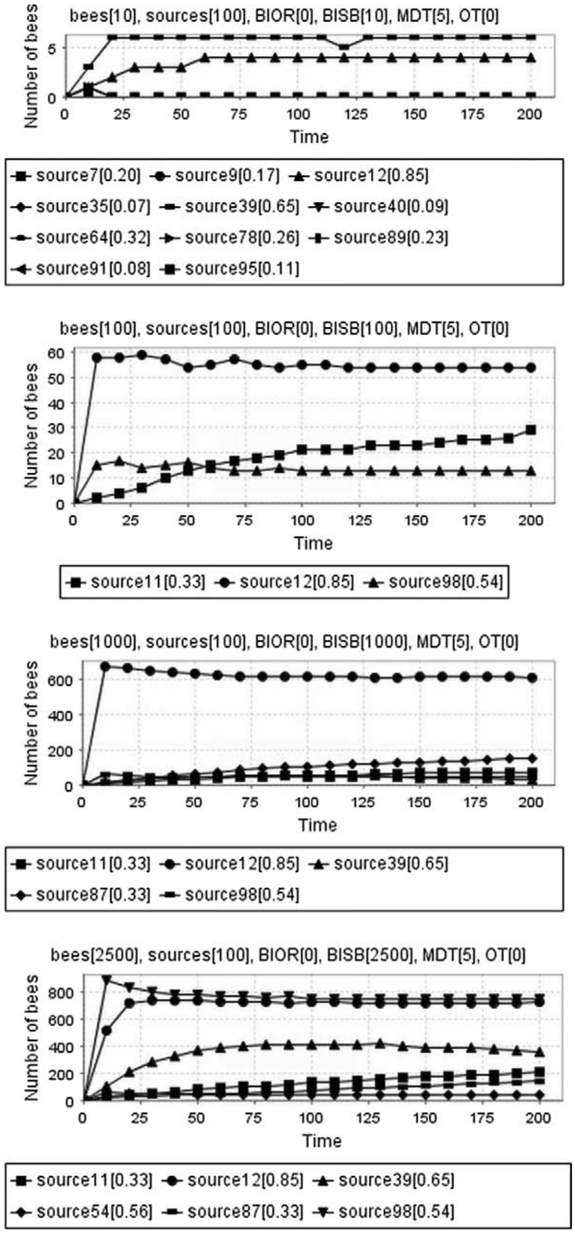

Fig. 5. Various numbers of bees forage for 100 sources. Sources have 3 attributes, a bee evaluates only one of them.

tion in the case of 2500 bees. They clearly failed in the case of 10 bees, which however should be attributed more to the low number of foraging bees than to the way of evaluating attributes. Still, results of this experiment cannot be interpreted in any conclusive way.

We developed a tool that implements our model so that we are able to perform various simulation experiments. In Fig. 6, there is depicted a user interface of our tool. The simulation tool allows to experiment with

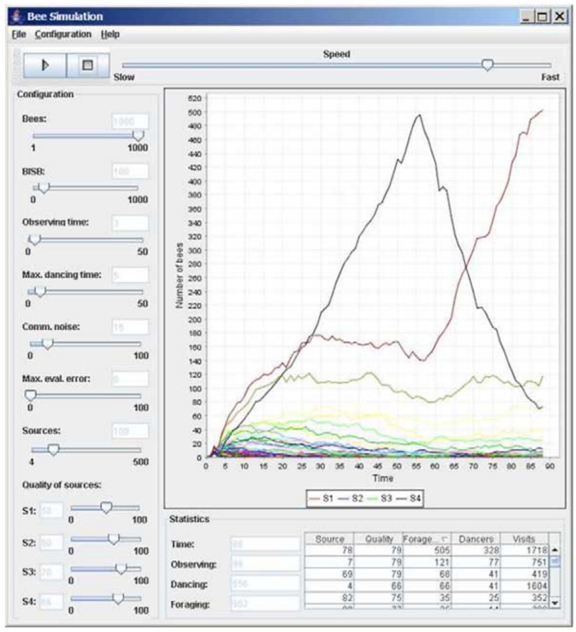

Fig. 6. Tool for simulation experiments with our model – a user interface.

uncertainty too. There are several sources of uncertainty that can be considered in this context. Let us introduce two kinds of uncertainties into our model. First, we admit that a bee dancer is able to communicate location of a source with some uncertainty (noise) only. Second, we acknowledge that a forager is able to evaluate the quality of a source with some uncertainty (error) only.

7. Web pages evaluation

Relevance of particular web page is from the user's point of view subjective. Each user has different interests and knowledge. There are some algorithms, which find out importance of particular web page according to its location in the graph of web and their interconnection in it. In this graph the vertices are formed by web pages and edges by interconnections among these pages. PageRank algorithm (Brin, Page, 1998) can be considered as a simulation of imaginary user that accidentally chooses various links on web. After each click the user decides if he continues, e.g. if he clicks on links on that page. The probability requesting another random page is called damping factor d. Many studies discussed the problem of appropriate values of damping factor. Generally this factor acquires value approximately 0.85 (Brin, Page, 1998). Each web page has unique PageRank, which is not dependent on the user's demand (query). This means that PageRank does not express the relevance of the page considering the given demand (query). This algorithm is used by Google to determine the criteria for sorting the search results of the user's query.

This algorithm uses the structure of hypertext links as a kind of the page "recommendation". The page evaluation is calculated not merely from the number of links that lead to this page but is based also on the evaluation of those pages. According to PageRank algorithm the page evaluation can reach the maximum value 10 points. We assume that pages $B, C, D, \ldots$ link to page A. If $L(X)$ is the number of links going out from page X, the formula for PageRank calculation which assumes the damping factor d, will be

$$PR(A) = 1 - d + d(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots)$$

Some versions of this algorithm indicate also another formula

$$PR(A) = \frac{1-d}{N} + d(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots)$$

where N is the number of all documents in the collection.

Regarding these formulas PageRank calculation of the certain page is derived from the PageRank of the other pages.

General formula for PageRank calculation of the page p_i is

$$PR(p_i) = 1 - d + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

where p_i is the page for which the PageRank is computed, $M(p_i)$ is the set of pages which "links" to the page p_i and $L(p_j)$ is the number of outgoing links from the page p_j. To receive the value of PageRank that would be sufficiently accurate, a number of iterations is needed. In each iteration the PageRank of every page in the collection is recalculated. The number of iterations depends on the number of pages and on the complexity of interconnection among component pages.

Iterative approach to PageRank calculation does not consider "the importance" of graph vertices, what will be demonstrated also in the following example.

Using iterative approach PR of the page A is calculated at the beginning, followed by calculation of PR of pages $B, C, \ldots, M$. As the most advantageous seems to be the approach, where we calculate as the first the PR of the pages from which the other pages depend. This is however not possible in the case that the dependence of the pages is cyclic (what of course can happen). Let us assume for example, that we would like to calculate PR of page E. That means we have to know the values of PR of pages B, D, H and F. To calculate PR of pages H and F we need to be familiar with the PR of other pages. According to this algorithm it would be appropriate to proceed in such a way, so that according to the actual needs PR of all dependent pages would be calculated.

Iterative approach, however, processes one vertex after another independently from their interconnection.

Here we find the model of bee hive for the calculation of PageRank very useful. Let us assume the behaviour of the bees would be slightly modified. The bee would fly to the source E (page or vertex E), she would then calculate $PR(E)$, but at the same time she would know, that to calculate $PR(E)$ more precisely, the values of $PR(D)$, $PR(B)$, $PR(F)$ and $PR(H)$ must be known. If the source E would have a high quality, the bee would fly into dance floor and dance there. However, she would not dance for the source E, but for one of the sources D, B, F or H. The choice of the source would be random with even assignment of choice probability for each of them. In this way the bee would try to attract other bees to come to source, for which she needs to calculate PR. After finishing dancing, the bee would fly back to the source E, whereby if

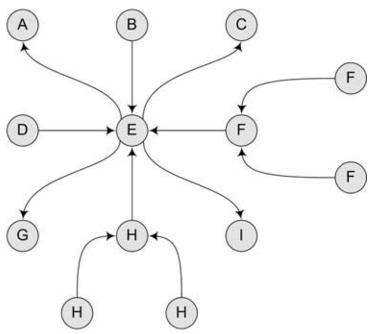

Fig. 7. The simple graph expressing the connections of web pages.

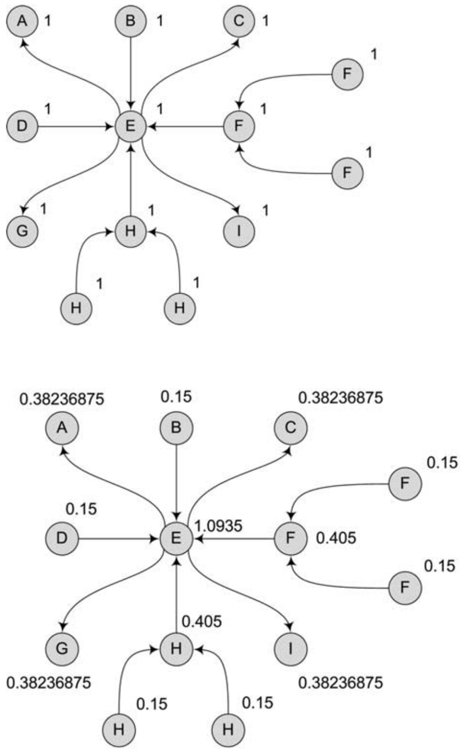

Fig. 8. Calculating PageRank of the simple graph.

she would succeed to attract the bee to fly to source D, B, F or H, the bee on the source E could determine $PR(E)$ more accurately. In the case that the quality of the source is high enough also after its actualization, there is a high possibility for the bee to fly to the dance floor again and dancing there for one of the sources D, B, F, or H. The whole cycle is repeated until the result value of $PR(E)$ is reached. Of course, similar rules hold valid for all sources. If the bee visits the source F and its quality is high enough, the bee will go to the dance floor and she will dance for one of the sources L or M.

After visiting the source s_i the bee calculates $PR(s_i)$ and consecutively determines the quality of source s_i as $Q_i = PR(s_i)/MaxPR$ where $MaxPR$ is the highest meanwhile known value of PR of some graph vertex. Further performance of the bee depends just on the calculated value of quality Q_i. The bee leaves the source s_i with the probability $P_X^{s_i} = 1 - Q_i$ and flies to auditorium. In the case the bee does not leave the source, she can in dependence from Q_{S_i} decide, if she flies to dance floor or she will continue in bringing the food from source s_i. If she flies to the dance floor, she will dance for one of the sources, which link to the source s_i. The choice of this source is random with even assignment of choice probability of any source linking to s_i. Let us call this source s_j. The bee will thus promote the source s_j. The bees in the auditorium always choose randomly one bee and they will follow her with the probability $P_F^{s_j}$. The probability $P_F^{s_j}$ is equal to the quality of source promoted by the bee in the dance floor $Q_j = PR(s_j)/PR(s_i)$.

In the case that the bee in auditorium will not choose any bee to follow after the given time limit, she flies to dispatch room, from where she will be sent to a random source. The cycle of graph vertices evaluation can be described as follows: At the starting time the value of each graph vertex is evaluated to $PR = 1$. The specified number of bees will accidentally start to fly out and visit the graph vertices. Every bee behaves after visiting the graph vertex (i.e. food source) according to described rules. This cycle is iteratively repeated until all the graph vertices are evaluated or when the given number of iterations is reached. The starting and final state of graph evaluation is depicted in Fig. 8.

8. Can bees evaluate pages?

Graph in Fig. 9 is very simple at a first view, but precisely the cyclic dependence causes that iterative PageRank computation is relatively time-consuming.

PageRank calculated by the standard approach has achieved the results after 52 iterations. The values of PageRank for the given vertices are in Table 1. The results of our model are in Table 2.

When using sufficient number of bees, good results are achieved. When employing 100 bees, only 13 iterations would be needed for this calculation. One of our further experiments was an experiment with the graph in Fig. 10. This graph has very dense interconnections of vertices with multiple cycles.

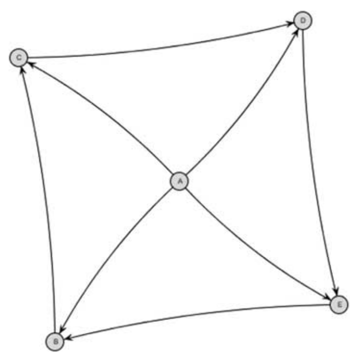

Fig. 9. Five vertices graph.

Table 1
Calculated PageRank

Vertex	Page Rank
A	0.15
B	1.2125
C	1.2125
D	1.2125
E	1.2125

Table 2
Number of iterations dependent on the number of bees

Number of iterations	Number of bees
13	100
20	50
40	25
80	10

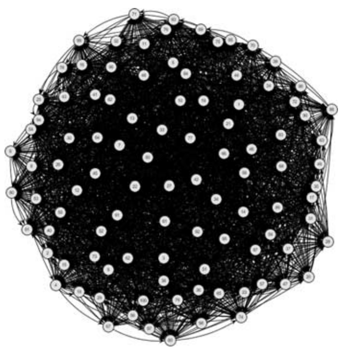

Fig. 10. Hundred vertices graph.

Table 3
Number of iterations dependent on the number of bees

Number of iterations	Number of bees
77	500
174	250
401	100
2011	50

As we can see from the results, when the number of bees is higher than the number of vertices in the graph, this method gives better results than the standard iterative method. But if we decrease the number of bees, number of iterations needed for PR calculation increases very rapidly.

When comparing in terms of accuracy of approximation, the higher number of bees has positive impact on the search results – it increases the possibility of visiting relevant sources. From the point of view of efficiency, our approach provides better results than conventional methods assuming that the number of bees is high enough (i.e. higher than the number of vertices in graph).

9. Conclusions and future work

A new approach to web search that is based on a bee hive metaphor was presented. Building on related works especially by Lorenzi et al. (2005), we proposed a modified model of a bee hive (first presented in (Navrat, 2006), further elaborated in (Navrat, Kovacik, 2006)). We have shown that the model is a true model of a bee hive in the sense it simulates several kinds of its typical behaviour Navrat et al. (2007). However, we have been even more keen to formulate a simple model that would describe some processes taking place in web search. We claim we have such a model, as some of our experiments show.

Our model still has basic properties characterising bees foraging for the food sources. Modifications have shown to be very useful especially from the point of view of decreasing the number of bees needed for searching high quality information.

Based on our results presented in (Navrat, Kovacik, 2006), we have investigated also the utilization of model of bee hive for the calculation of graph PageRank. In many cases computation on the basis of bee hive model requires fewer iterations than the standard iterative approach. In contrast to the iterative method, our method does not provide as accurate results, since we have no guarantee that the bees will visit every source. This property can be rather easily compensated by increasing the number of bees, thus decreasing the probability that the bees will not visit all the sources. It should be noted, however, that in a strict sense, comparing numbers of iterations does not replace a comprehensive complexity analysis. For example, details of communication overhead in terms of messages could come into play.

Another advantage of this approach is the possibility of distributed PageRank calculation of a large graph. It should not introduce any substantial difficulty to distribute bees access more computers, where they would communicate with each other also in the case they would be located on different computers. In this way would bees be usable also for the computation of PageRank of huge graphs. On the other hand, we are aware of the possible problem of scalability of our approach and further research is clearly needed in this direction.

We see one more advantage, which is the possibility of continual computation of graph PageRank on-line. For instance Google recalculates PageRanks of all indexed processed pages once a certain period of time. If we would use the bees, it would be relatively easy to recalculate PageRanks continuously – i.e. the bees would constantly fly over the graph and they would update PageRank of individual vertices (web pages) of graph.

We are well aware more research is still ahead. However, the bee hive metaphor has received further endorsement. At the same time, to reach a workable model of web search requires not only refinement, but more ideas. The idea of bee typing according to types of attributes they evaluate is to be elaborated. We have not attempted to introduce any schemas how to combine partial results, although there are obvious candidates such as weighting etc. The model itself can be more refined. Additional parameters, such as initial distribution of foraging directions could be considered. Alternatively to generating initial direction randomly, some kind of memory (individual or collective) could be considered helping to retrieve directions to sources that have been useful in the past in providing information relevant to the current or similar past query.

Some reflection on the very purpose of the model might be desirable, too. We have used the term web search in a rather loose sense. It appears as retrieving information from the web could benefit from this approach. In particular, information recommending and filtering could be modelled by the bees' behaviour. But other possibilities are still open for research.

The bee hive metaphor has been used so far as a basis for a model of recommending information from a given set of sources collected from the web. Besides such an off line application, searching a "frozen" part of the web, it might be worthwhile to attempt to search the "living" web with bees becoming agents. Such a dynamic search, if performed continuously, could open possibility of adaptive recommendation according to changes (disappearances of old documents and appearances of new ones) on the web.

Acknowledgements

This work was partially supported by the Slovak State Programme of Research and Development "Establishing of Information Society" under the contract No. 1025/04 and the Scientific Grant Agency of Republic of Slovakia, grant No. VG1/3102/06.

References

A. Abraham and V. Ramos, Web usage mining using artificial ant colony clustering and genetic programming, in: *CEC'03 – Congress on Evolutionary Computation*, Canberra, Australia, 8-12 Dec. 2003, IEEE Press, ISBN 0-78037-804-0, pp. 1384–1391.

S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, Computer Science Department, Stanford University, Stanford, http://www-db.stanford.edu/˜backrub/google.html.

H.E. Bullock, P. Dey, and K.D. Reilly, A "Bee Hive" model for heterogeneous knowledge in expert systems, *ACM*, 1986, pp. 417–417.

S. Camazine and J. Sneyd, A model of collective nectar source selection by honey bees: Self organization through simple rules, *Journal of Theoretical Biology* **149**(4) (1991), 547–571.

D. Chuda, Visualization in education of theoretical computer science, in: *Proc. Int. Conference on Computer Systems and Technologies – CompSysTech'07*, B. Rachev, A. Smrikarov, D. Dimov, eds, Rousse, Bulgaria, 2007, pp. IV-15-1–IV-15-6.

D. Chuda, Evaluation and security features in e-learning, in: *e-learning Conference'07*, A. Gungor, N. Aydin, A. Karahoca, C. Toklu, eds, Istanbul, Turkey, 2007, pp. 81–85.

H. de Vries and J.C. Biesmeijer, Modelling collective foraging by means of individual behaviour rules in honey-bees, *Behav. Ecol. Sociobiol.* (1998), 109–124.

A. Dornhaus, F. Klugl, F. Puppe, and J. Tautz, Task selection in honeybees – experiments using multi-agent simulation, in: *3rd German Workshop on Artificial Life*, 1998, pp. 171–183.

D. Gordon, The organization of work in social insect colonies, *Nature* **380** (1996), 121–124.

F. Heylighen, Collective intelligence and its implementation on the web: algorithms to develop a collective mental map, *Computational & Mathematical Organization Theory* **5**(3) (1999), 253–280.

L. Hluchy, M. Laclavik, Z. Balogh, and M. Babik, AgentOWL: Semantic knowledge model and agent architecture, *Computing and Informatics* **25**(5) (2006), 421–439.

N. Holden and A.A. Freitas, Web page classification with an ant colony algorithm, in: *Parallel Problem Solving from Nature – PPSN VIII*, Xin Yao et al., ed, LNCS 3242, Springer-Verlag, September 2004, pp. 1092–1102

F. Kluegl, C. Oechslain, F. Puppe, and A. Dornhaus, Multi-agent modelling in comparison to standard modelling, in: *Proc. AIS2002 Artificial Intelligence, Simpulation and Planning in High Autonomy Systems*, F.J. Barros, N. Giambasi, eds, SCS Publ. House, 2002, pp. 105–110.

R. Koval and P. Navrat, Intelligent support for information retrieval of web documents, *Computing and Informatics* **21**(5) (2002), 509–528.

H. Lei and V. Govindaraju, Matching and retrieving sequential patterns using regression, *Web Intelligence and Agent Systems* **3**(4) (2005), 261–270.

F. Lorenzi, D. Scherer dos Santos, and A.L.C. Bazzan, Negotiation fopr task allocation among agents in case/base recommender systems: a swarm intelligence approach, in: *Proc. IJCAI 2005 Conference, Workshop*, 2005, pp. 23–27.

F. Lorenzi, D. Scherer dos Santos, and A.L.C. Bazzan, Case-Based Recommender System Inspired by Social Insects, in: *Proc. XXV Congresso da Sociedade Braseilera de Computacao*, Sao Leopoldo, 2005, pp. 752–760.

K. Matusikova and M. Bielikova, Social navigation for semantic web applications using space maps, *Computing and Informatics* **26**(3) (2007), 281–299.

P. Navrat, Bee hive metaphor for web search, in: *CompSysTech 2006*, B. Rachev, A. Smrikarov, eds, Veliko Turnovo, Bulgaria, June 2006.

P. Navrat and M. Kovacik, Web search engine as a bee hive, in: *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06)*, Hong Kong, 2006, pp. 694–701.

P. Navrat, M. Kovacik, A. Bou Ezzeddine, and V. Rozinajova, Information retrieval using honey bees, in: *Znalosti 2007*, Ostrava, 2007, pp. 63–74.

S.L. Ong, W.K. Lai, T.S.Y. Tai, and K.M. Hoe, Application of ant-based template matching for web documents categorization, *Informatica* **29** (2005), 173–181.

S.J. Schultze, A collaborative foraging approach to web browsing enrichment, in: *Proc. CHI 2002*, ACM, 2002, pp. 860–861.

T.D. Seeley, S. Camazine, and J. Sneyd, Collective decision-making in honey bees: how colonies choose among nectar sources, *Behav Ecol Sociobiol* **28** (1991), 277–290.

C.A. Tovey, The honey bee algorithm. A biologically inspired approach to internet server optimization, *Engineering Enterprise* (Spring 2004), 13–15.

K. Visscher and S. Camazine, Collective decisions and cognition in bees, *Nature* **397** (1999), 400–400.

S. Ye, T. Chua, and J. Kei, Clustering web pages about persons and organizations, *Web Intelligence and Agent Systems* **3**(4) (2005), 203–216.

X. Zhang, V. Lesser, and T. Wagner, A Layered Approach to Complex Negotiations, *Web Intelligence and Agent Systems* **2**(2) (2004), 91–104.