

# Bee Hive at Work: Following a Developing Story on the Web

Pavol Navrat and Anna Bou Ezzeddine

Faculty of Informatics and Information Technologies, Slovak University of Technology  
Bratislava, Slovakia  
navrat@fiit.stuba.sk, ezzeddine@fiit.stuba.sk

**Abstract.** Problem solving methods, inspired biologically or by nature, are gaining in importance. Their potential for devising new, but also for increasing efficiency of some well-known methods for solving different kinds of problems, is becoming more and more obvious. The behaviour of social insects provides a lot of inspiration. We took inspiration from honey bees, in particular the way they behave in a hive. We proposed a modified model devised for searching parts of the web online. In a series of experiments, we attempted to follow a developing story with the aim to automatically find documents (web pages) bringing news on the story. This service, working as a beehive, delivered documents as expected.

**Keywords:** bee hive model, bee swarm intelligence, web search, on-line web crawling, developing story.

## 1 Introduction

When seeking new ways how to solve various classes of difficult problems, researchers more and more often seek inspiration from several animal species, observing as they behave in nature. At first sight, behaviour of particular animals may seem primitive, but when they act in a swarm, they can achieve states that can be interpreted as solutions to quite complicated problems. Moreover, the problem solving process is often quite effective. Processes such as collecting food by ants or foraging honey by honey bees are an instant inspiration. From observing them, simple rules can be inferred. Rules constitute basis for algorithms to solve various classes of interesting problems.

Among the animal species that have been attracting the interest of researchers, social insects occupy a dominant position. It is especially so because of the simple behaviour of an individual in relation to a complicated collective behaviour. A society (or swarm) of social insects is able to evolve means for collective problem solving. Complexity of such problems is well above the abilities of an individual. They can solve problems without central management and without predetermined fixed structures despite the massive internal whirr.

At first sight different kinds of social insects, such as ants, honey bees, wasps or termites, have different kinds behaviour. However if we study their behaviour and activities in greater detail, we can find many similar principles.

The behaviour of honey bees is a study subject of researchers in various disciplines, mainly biologists and social biologists in particular.

Their studies have proven to be extremely useful [1],[3],[4],[5],[24],[27]. They provide the underlying information on how honey bees behave in nature. We dare to identify one additional new dimension for problem solving: the web. It brings new opportunities, since it poses new kinds of problems. Nowadays we have a great wealth of information instantly retrievable. Therefore, we identify web problem solving as another excellent area of research.

The area is of course very broad. We have chosen one interesting information gathering scheme, which has become quite typical nowadays, when news appear on the web continuously. An event is reported. However, it induces other events and reactions, so in effect a story is developing. All this leaves traces on web pages of media portals reporting on it. We attempt to tackle this problem, moreover, we attempt to tackle it genuinely on-line. This is a true challenge for a bee hive.

The rest of the paper is organized as follows. In Section 2, we briefly describe the concept of self-organization in biological systems. We continue with briefly mentioning related works, then in Section 4 we describe the bee hive model and in Section 5 the web search it performs, concentrating on explaining the way a bee evaluates quality of a source. In Section 6, we elaborate how our bee hive tracks a story on the web. Section 7 concludes and hints future work.

## 2 Self-organization in Biological Systems

Intelligent behaviour of social insects begins through indirect communication among the individuals of the community known as the stigmergy principle.

The term of stigmergy was coined in the late fifties of the last century [7], to denote types of correlation among social insects that arose from for example, building mounds by termites and ants, or hives by bees. Stigmergy is a mechanism of spontaneous indirect coordination among the individuals in a community. It is a form of self-organisation, in which the community forms complex, declaredly intelligent, structures without the necessity of either planning or management. It supports effective coordination among very simple animals, that do not possess any memory or intelligence whatsoever.

This term was later transferred also to other fields including computer science. The term swarm intelligence was according to [3] used in artificial intelligence in 1989 by mobile robot systems.

Swarm intelligence describes collective behaviour in a decentralized self-organised system that can be natural or artificial. Swarm intelligence, in terms of artificial intelligence, represents the ability of systems composed of non-intelligent agents with limited individual abilities, to achieve general intelligent behaviour of the conjuncted swarm [26]. The agent is at this time known as an individual that is able to sense and observe its surroundings and consequently undertake some of the available actions. These actions also include changes in the surroundings that are occupied by the agent. Individuals do not possess information about how to solve the existing problem and

the intelligent behaviour appears as a consequence of actions of the swarm represented by these agents.

### 3 Related Work

Description of self organization of a honey bee system is discussed in [6]. Simple rules help the bees select the best source of nectar. Nectar is essential for their survival, so they fly out to their surroundings and look for it. Finding a food source compels the bee to fly back to the bee hive and bring the information about the source.

Models of bees collectively looking for collecting nectar are dissertated in [25]. Individually oriented simulation is constructed to simulate collective behaviour of bees in time of foraging. Each bee has to follow the same set of behavioural rules. Building up a simulation model, which would achieve similar results as those presented in [24], has been the main goal.

Camazine in [5] presents a description of dynamical interaction among bees in the process of carrying nectar from two sources using mathematical simulation.

The mathematical model of [5] has been an inspiration for a prototype of multi-agent recommending system, which was proposed in [10] and presented in [11] and [12]. They show the way of utilizing the bee hive metaphor. They drew from works of [23] and [22] applying different approaches to solving problems. Lorenzi worked in [9] on this question as well. By allowing more than two sources of food, they generalised the model of [5]. The foible of their model is that it presupposes the existence of as many bees as there are sources.

### 4 Bee Hive Model

In the process of devising the model, we inspired ourselves by a model introduced in [12]. Some limitations of the model were identified so the main aim was to eliminate them.

We considered as perhaps the main limitation of the model [12] that it included the initial assignment of one honey bee to one food source. Using this type of assignment for achieving valid results requires using as many honey bees – agents - as there are sources. Existence of a greater number of sources (e.g. web pages) incapacitates the model.

Parameters of the model are:

- total number of honey bees in the model  $N$ ,
- maximum dancing time for a particular food source  $MDT$ ,
- maximum time that the honey bee spends in the auditorium  $OT$ ,
- information whirr  $NOICE$ , accuracy in exchanging information between the dancing bee and the observing bee,
- error in evaluating the quality of the source  $ERR$ , its value is from the interval  $\langle 0,1 \rangle$ .

Our model consists of three parts: dispatch room, dance floor and auditorium.

Every honey bee starts from the dispatch room, chooses a source randomly, evaluates its quality  $q$  and returns back to the hive. There it decides whether it remains with the discovered source or moves to the auditorium to observe dancing bees.

If it decides to remain with the source, it will get to the dancing room with the probability  $q$  to promote the given source, or it will return to its source with the probability  $(1-q)$ .

The length of the dance time in the dancing room is proportional to the quality of the source, but it cannot be greater than MDT. After terminating its dance the bee returns back to its source.

If the bee decides to abandon its source, it enters to the auditorium, where it chooses to watch some of the dancing bees randomly. With the probability that is assigned by the ratio of the total number of bees dancing for the same source as the selected bee and the number of all bees in the dancing room, the observing bee would follow the recommended source. Otherwise it would remain in the dancing room and attempt to choose a different source. The stay of the bee in the auditorium must not overpass the value OT. The bee proceeds to the dispatch room after the expiration of the OT period.

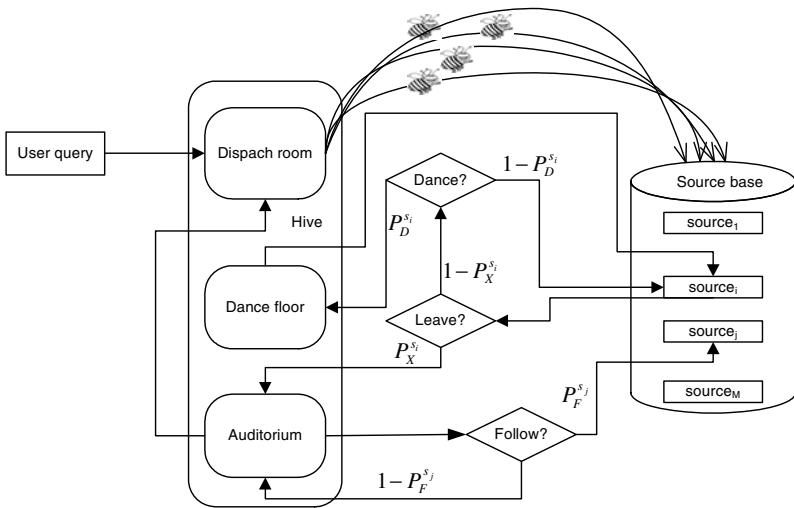


Fig. 1. Behaviour model of honey bee community according to [14]

The mechanism of our model can be found in [14], which contains a more detailed elaboration. In [15] and [17] we present additional experiments with the parameters of the model. The original model is universal and it can be used in various domain fields. It defines the behaviour of honey bees in the hive. In [16] we explicated the behaviour of bees when being outside the hive for Internet search.

In the process of searching on the Internet, the web page is considered to be the source, and the URL address is regarded as its location.

## 5 Web Search

In [8], different methods of honey bee swarm utilization in information systems are described in greater detail. As we mentioned, the priority of our research is web search, in particular on-line search inspired by the behaviour of a swarm.

Nowadays majority of locators work on off-line database, which includes indexed web pages from some parts of the Internet. The contents of these web pages are updated constantly and new web pages are accumulated.

The volume of Internet is enormous therefore it is impossible to search through all web pages. This is why it is necessary to focus the web search only on the most promising way to search. Every user possesses some preferences when searching the web. They can be expressed by selecting a cluster of preferred web pages. Web search begins from a selected cluster of web pages. So a web page represents a source and the accessibility of another source represents hypertext reference to another web page (source).

One of the most important functions of the model is detecting the quality of the found web page. Quality of the source is from interval  $\langle 0,1 \rangle$ . The resultant quality is made of partial qualities. In particular cases it is possible to select partial qualities and their importance in the process of determining the whole quality of the source.

Partial qualities:

1. Quality of the range between two web pages – it is defined as a number of different domains that a bee visits in order to get from one web page to another.

2. Incidence number quality – quality based on the number of word incidence. It is computed in agreement with the formula:

$$q_{count} = \frac{-1}{2 \left( n + \frac{1}{2Q_{COUNT}} \right)} + Q_{COUNT} \quad (1)$$

where  $n$  is the number of incidence of the given word and  $Q_{COUNT}$  is the maximal value defined for this partial quality.

3. Quality of incidence in the caption – in this case it is important to trace the hierarchical level of word incidence in captions. Numbers from 0 to 6 are associated to every level. Calculation of this partial quality is given by formula:

$$q_{header} = Q_{HEADER} - h * \frac{Q_{HEADER}}{HEADER_{MAX} + 1} \quad (2)$$

where  $h$  represents the minimum of all caption values that include the searched word;  $Q_{HEADER}$  is the maximum value of this partial quality;  $HEADER_{MAX}$  is the maximal depth of the caption hierarchy that the calculation takes into account.

4. Readability index by Flesch – this test evaluates the text according to a 100 degree scale. The higher the score is, the more comprehensible the text. The formula for Flesch's readability index is:

$$206.835 - (1.015 \times ASL) - (84.6 \times ASW) \quad (3)$$

where ASL = Average Sentence Length (the number of words divided by the number of sentences), ASW = Average number of Syllable per Word (the number of syllables divided by the number of words). In [18] we have shown that using the combination of key words and readability index makes the process of obtaining promising results possible, especially in professional text search.

## 6 Web Story

When an inquisitive person (IP), a curious fellow, takes interest about a particular event, which is being monitored by media, he or she wants to read news reporting on that event. Often, an event is not an isolated one. There is some follow up, some reactions, induced events. A story develops. IP wants to follow the story. Sometimes, the news itself is the most important information. But usually, reports are soon followed by commentaries, in-depth analyses, blogs or discussions. IP may be interested in some of this as well. It is fair to assume all this data is available on the web. To access just the right documents requires searching, filtering, sorting and other data processing, which is time-consuming. Moreover, since the story develops in time, it would require almost continuous monitoring of the story as it progresses.

Thus our aim was to devise a method implementable on a personal computer that would be capable of supplying documents related to the developing story as they are emerging on the web for several hours or days.

It seems that any such method must include searching the web. After collecting a set of related documents, it is necessary to filter and cluster the data set. News portals contain mainly articles, annotations, discussions, blogs and symposia. Articles are important from the aspect of information tracing. Annotations, blogs and symposia are elaborated in order to gain opinions on particular events, eventually on offered products [21].

Information that was found and sorted was classified according to the publication date. We attempted to approach this problem using the bee hive model [19]. Inspired by the work of [13] we used the concept of energy from agent InfoSpider.

Source quality is calculated as the sum of partial qualities (incidence number quality, caption incidence quality, Flesch's readability index ).

Biological observations show [1] that bees are specialized as either scouts or recruits in the process of collecting food. Scouts look for food independently; recruits gather food with the use of information from other bees.

In our model a bee in the auditorium can be considered as a free bee. If it chooses a bee from the dancing room and will follow that bee's source, it becomes a recruit.

If it does not choose a bee and passes to the dispatch room, it becomes a scout. Scouts and recruits act similarly when being inside the hive, but their behaviour changes as soon as they leave the hive. After finding a source with non zero quality, the scout returns to the hive to promote its source.

Parameters of our model of the hive are given in Table 1, and of the quality calculation in Table 2.

**Table 1.** Parameters of the hive

Number of bees	30
MDT maximal dancing time	7 iteration
OT maximal time in auditorium	4 iteration

**Table 2.** Parameters of quality calculation

Initial energy of a bee	1
Energy increase	source quality
Energy decrease because of transition to new source	0.05
Maximum quality of incidence number	0.7
Maximum caption incidence quality	0.15
Maximum caption number	3
Flesch's readability index	0.15

## Experiments

We used the event of earthquake in Haiti that has been widely monitored by media, at the time as a developing story. In our experiment, three news portals represent the start web pages: [www.pravda.sk](http://www.pravda.sk), [www.sme.sk](http://www.sme.sk) and [www.ta3.com](http://www.ta3.com). The key words we looked for were: earthquake and Haiti.

The process of following the story was divided into three parts.

The first part is represented by an experiment that took place from 13 January 2010 10:00 am to 14 January 2010 4:00 pm 9327 web pages were found, out of which 1066 were of non zero quality. Web pages of non zero quality were divided into 5 classes:

**Table 3.** Division of related pages into classes, 13-14 January 2010

Informative page	493
List of articles	348
Discussions	73
Blogs	65
Graphic content (pictures, videos)	87

Let us assume the IP seeks new raw information on this story, that has been developing so dramatically. For the IP we have in mind, discussions and blogs are irrelevant because they represent only reactions to the event.

The list of articles has no informative value, but is important for page discovery.

Because of sufficient number of informative web pages, we selected only web pages with quality higher than 60 percent. (There were 217 such web pages). These web pages were classified according to the published date extracted from the page.

The most frequently used words on those 217 informative web pages were: disaster, tragedy, victims, UN, chaos, help.

The second part of the experiment took place on 16 and 17 January 2010 always at the same time: from 8:00 am to 5:00 pm 11439 web pages were found, 1193 with non zero quality. Web pages of non zero quality were divided into these classes:

**Table 4.** Division of related pages into classes, 16-17 January 2010

Informative page	552
List of articles	385
Discussions	83
Blogs	72
Graphic content (pictures, videos)	101

298 informative web pages with the quality higher than 60 percent were found.

The most frequently used words which occurred on the web pages after 16 January 2010 8:00 am were: cadavers, indigence, looting, despair, water, help.

The last part of the experiment took place from 18 to 21 January 2010, always at the time from 6:00 pm to 11:00 pm and 12576 web pages were found, out of which 1271 of non zero quality. Web pages of non zero quality were divided into these classes:

**Table 5.** Division of related pages into classes, 18-21 January 2010

Informative page	598
List of articles	399
Discussions	87
Blogs	78
Graphic content (pictures, videos)	109

327 informative web pages possessing a quality of above 60 percent were found.

The most frequently used words on the web pages after 18 January 2010 6:00 pm were: water, collections, help, charity, putrefaction, physicians.

Due to the results of the experiments we can deduce, that our designed system is able to follow a story as it develops. When repeating the experiment twice or three times, the algorithm found almost all the web pages that were marked as relevant by the previous algorithm run, hence supporting a hypothesis that our method based on the modified bee hive model is quite robust.

## 7 Conclusion and Future Work

The bee hive model has been used in our previous works mainly in web search and in function optimisation [20]. In this work we present a modified bee hive model and employ it in a system for tracking a developing story.

From the experiments we can conclude the following:

- the system is able to look for web pages and evaluate the quality of the found web pages automatically,
- it can collect relevant pages,
- it can reconstruct the story backwards in time,



- it can monitor the story that is developed during the search,
- it provides statistical results about the searching process,
- by means of this system we can obtain the most frequently used words on time distinguishable web pages.

According to the acquired statistical results we can conclude that after the initial wide interest of the population about the mentioned disaster (the first part of the experiment), the interest stabilizes gradually. Number of discussions and blogs did not change largely through the whole duration of the experiment.

Due to the obtained key words we can deduce the changing contents of the relevant articles.

In the forthcoming period we plan to process the acquired discussions and blogs and to evaluate the opinions of discussants on possible ways of help in disasters.

## Acknowledgement

This work was partially supported by the grants VEGA 1/0508/09, KEGA 345-032STU-4/2010 and it is the partial result of the Research & Development Operational Programme for the project Support of Center of Excellence for Smart Technologies, Systems and Services, ITMS 26240120029, co-funded by the ERDF.

## References

1. Biesmeijer, J.C., de Vries, H.: Exploration and exploitation of food sources by social insect colonies: A revision of the scout-recruit concept. *Behavioral Ecology and Sociobiology* 49(2-3), 89–99 (2001)
2. Bonabeau, E., Theraulaz, G., Deneubourg, J.L., Aron, S., Camazine, S.: Self-organization in social insects. *Trends in Ecology and Evolution* 12, 188–193 (1997)
3. Bonabeau, E., Thraulaz, G., Dorigo, M.: *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, Oxford (1999)
4. Bullock, H.E., Dey, P., Reilly, K.D.: A “Bee Hive” Model For Heterogeneous Knowledge. In: *Expert Systems*, pp. 417–417. ACM, New York (1986)
5. Camazine, S., Sneyd, J.: A model of collective nectar source selection by honey bees: Self-organization through simple rules. *Journal of Theoretical Biology* 149(4), 547–571 (1991)
6. Camazine, S., Deneubourg, J.L., Franks, N.R., et al.: *Self-organization in Biological Systems*. Princeton University Press, Princeton (2003)
7. Grasse, P.-P.: La Reconstruction du nid et les Coordinations Individuelles chez *Belligositermes Natalensis* et *Cubitermes* sp. la Theorie de la Stigmergie: Essai d’interpretation du Comportement des Termites Constructeurs. *Insectes Sociaux* 6, 41–80 (1959)
8. Karaboga, D., Akay, B.: A survey: algorithms simulating bee swarm intelligence. *ArtifIntellRev*, SpringerScience Busines Media B 31(1-4), 61–85 (2009)
9. Lorenzi, F., Ricci, F.: Case-based reasoning and recommender systems. Technical report, IRST (2004)
10. Lorenzi, F., Sherer Dos Santos, D., Bazzan, A.L.C.: Case-based Recommender Systems: a unifying view. In: *LNCS/LNAI State-of-the-Art Survey book on Intelligent Techniques in Web Personalization* (2005)

11. Lorenzi, F., Sherer Dos Santos, D., Bazzan, A.L.C.: Case-based recommender systems inspired by social insects. In: XXV Congresso da Sociedade Brasileira de Computacao, July 22-29 (2005)
12. Lorenzi, F., Sherer Dos Santos, D., Bazzan, A.L.C.: Negotiation for task allocation among agents in case-based recommender systems: a swarm-intelligence approach. In: IJCAI 2005 Workshop on Multi-Agent Information Retrieval and Recommender Systems, Edinburgh, pp. 23–27 (2005)
13. Menczer, F., Monge, A.E.: Scalable Web Search by Adaptive Online Agents: An InfoSpiders Case Study. In: Intelligent Information Agents. Springer, Heidelberg (1999)
14. Navrat, P., Kováčik, M.: Web Search Engine as a Bee Hive. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 694–701. IEEE Computer Society, Los Alamitos (2006)
15. Navrat, P., Kovacik, M., Bou Ezzeddine, A., Rozinajova, V.: Information retrieval using honey bees. In: Znalosti 2007, Ostrava, pp. 63–74 (2007)
16. Navrat, P., Jastrzemska, L., Jelínek, T., Bou Ezzeddine, A., Rozinajová, V.: Exploring Social Behaviour of Honey Bees Searching on the Web. In: Li, Y., Raghavan, V.V., Broder, A., Ho, H. (eds.) 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (Workshops), Silicon Valley, USA, November 2-5, pp. 21–25. IEEE Computer Society, Los Alamitos (2007)
17. Navrat, P., Kováčik, M., Bou Ezzeddine, A., Rozinajová, V.: Web search engine working as a bee hive. In: Web Intelligence and Agent Systems: An International Journal, vol. 6, pp. 441–452. IOS Press, Amsterdam (2008)
18. Navrat, P., Taraba, T., Bou Ezzeddine, A., Chudá, D.: Context Search Enhanced by Readability Index. In: Artificial Intelligence in Theory and Practice II. IFIP 20th World Computer Congress. IFIP Series, vol. 276, pp. 373–382. Springer Science+Business Media, LLC, New York (2008)
19. Navrat, P., Jastrzemska, L., Jelinek, T.: Bee Hive At Work: Story Tracking Case Study, Milan, Italy, September 15-18, pp. 117–120. IEEE Computer Society, Los Alamitos (2009)
20. Navrat, P., Jelinek, T., Jastrzemska, L.: Bee Hive At Work: A Problem Solving, Optimizing Mechanism. In: World Congress on Nature and Biologically Inspired Computing (NaBIC 2009) & 8th International Conference on Computer Information Systems and Industrial Management Applications (CISIM2009), December 9-11. PSG College of Technology, Coimbatore (2009)
21. Navrat, P., Bou Ezzeddine, A., Slizik, L.: Mining Overall Sentiment in Large Sets of Opinions, AWIC 2009. In: Advances in Intelligent Web Mastering - 2. AISC, vol. 67, pp. 167–173. Springer, Heidelberg (2010)
22. Resnick, P., et al.: An open architecture for collaborative filtering of netnews. In: Proceedings ACM Conference on Computer-Supported Cooperative Work, pp. 175–186 (1994)
23. Schafer, J.B., Konstan, J.A., Riedl, J.: E-commerce recommendation applications. *Data Mining and Knowledge Discovery* 5(1/2), 115–153 (2001)
24. Selley, D., Camazine, S., Sneyd, J.: Collective decision-making in honey bees: How colonies choose nectar sources. *Behavioral Ecology and Sociobiology* 28, 277–290 (1991)
25. Vries, H., Biesmeijer, J.C.: Modelling collective foraging by means of individual behaviour rules in honey-bees. *Behav. Ecol. Sociobiol.* 44, 109–124 (1998)
26. White, T., Paturek, B.: Towards Multi-Swarm Problem Solving in Networks. In: Multi Agent Systems - International Conference on Volume (1998)
27. Zhang, S., Schwarz, S., Pahl, M., Zhu, H., Tautz, J.: Honeybee Memory: A Honeybee Knows What To Do And When. *Journal of Experimental Biology* 209, 4420–4428 (2006)