Towards a Collaborative Ranking Mechanism for Efficient and Personalized Internet Search Service Provisioning

Malamati Louta[1]✉, Ioannis Anagnostopoulos[2], Ilias Maglogiannis[2], Angelos Michalas[3]

[1]Technological Educational Institute of Western Macedonia, Department of Business Administration
[1]University of Western Macedonia, Department of Information and Communication Technologies Engineering
GR-50100, Kozani, Greece
e-mail: louta@telecom.ntua.gr

[2]University of Aegean, Department of Information and Communication Systems Engineering
Karlovassi, GR-83200, Samos Island, Greece
e-mail: {janag,imaglo@aegean.gr}

[3]Technological Educational Institute of Western Macedonia, Department of Informatics and Computer Technology
GR-52100, Kastoria, Greece
e-mail: amichalas@kastoria.teikoz.gr

Abstract

The aim of this paper is, in accordance with efficient web search service operation objectives, to propose enhancements to the sophistication of the functionality that can be offered by search engine services. A meta-search third-party result ranking mechanism is proposed, which is capable of adapting over the continuous changes that occur on the web, providing in parallel personalized information acquisition considering the user's navigation behaviour. Transparency is achieved for both personalization and web evolution adaptation mechanisms, requiring virtually none effort from the user's part. In essence, the proposed meta search engine rates, re-organises and combines the results acquired from search services for a specific user information resource request in accordance with a weighted combination of a performance related factor (tightly related to the ranking of the web result as given by the search engine service) and a reliability related factor (corresponding to the user satisfaction stemming from the specific web result that he/she browses), while the performance of each search engine with respect to adequately adapting to the web evolution is taken into account. For the evaluation of the web results reliability, a collaborative reputation mechanism is utilized, which helps estimating their quality and predicting their future usability, taking into account their past performance in consistently satisfying user expectations. A set of results indicative of the efficiency of our proposed scheme is provided. The Internet search services used were Google, MSN and Yahoo!

✉ Corresponding Author

Key Words: Meta-search algorithm, personalized information acquisition, web evolution adaptation, reputation mechanism, performance and reliability related factors.

1 INTRODUCTION

The vast increase of web resources has boosted the demand for effective personalized information resources search and acquisition. In this perspective, web search engine services have a vital role, since they form an information broker between the user and the huge amounts of disseminated information. Considering the fact that (in most cases) it is difficult for the users to adequately and/or accurately describe their requirements and constraints with keywords, the search services return a vast amount of results, presenting lower precision in the first recall levels (top-ranked results). Thus, the construction of user profiles for personalized information search is necessitated.

Related work in the research literature involves personalized search techniques based on user profiling [1], [2]. In these techniques, the users are asked to fill forms describing their interests, or they are asked to label their information needs among already built categories and taxonomies [3], [4], [5]. Taking into account that most of the users are unwilling to provide explicit feedback on either their interests or the returned results (such interaction is considered as an additional overhead in the users browsing activity as shown by studies conducted in the field of human-computer interaction [6]), the task of automatically building user profiles capable of adapting to the users' interests is quite challenging. To this respect many approaches have been proposed and discussed (e.g., [7], [8], [9], [10]), in which the user's interaction with the browser and his/her respective patterns of behaviour are implicitly recorded and evaluated. Additionally, search services should adequately adapt to the web evolution in a way that new information is indexed as fast as possible, dead links are removed and the validity of the results is checked frequently [11], [12], [13].

The subject of this study falls into the overall search service provisioning procedure, trying to extend pertinent previous work in the literature. Specifically, the aim of this paper is, in accordance with efficient Internet search service operation objectives, to propose enhancements to the sophistication of the functionality that can be offered by search engine services. A *meta-search third-party result ranking mechanism* is proposed, which is capable of adapting over the continuous changes that occur on the web, providing in parallel personalized information with respect to the user's navigation behaviour. Transparency is achieved for both personalisation and web evolution adaptation mechanisms, requiring virtually none effort from the user's part.

In essence, the proposed meta search engine rates, re-organises and combines the results acquired from search services for a specific user request in accordance with a weighted combination of a *performance related factor* (tightly related to the ranking of the web result as given by the search engine part) and a *reliability related factor* (corresponding to the user satisfaction stemming from the investigated web result). For the evaluation of the web results reliability, a *collaborative reputation mechanism* is proposed, which helps estimating their quality and predicting their future usability, taking into account their past performance in consistently satisfying user expectations. The reliability criterion is motivated by the fact that there may be different levels of user satisfaction with respect to the disseminated content of each web result. In this respect, there may be web results that, in principle, do not meet user requirements and preferences. Hence, recording the previous experience can easily assist the meta-search engine in deciding how to present to the user the results obtained from the search services. User satisfaction evaluation requires for a *browsing behaviour monitoring mechanism*, which records user's navigation behaviour during search sessions in a fully transparent way. Additionally, the performance of each search engine with respect to adequately adapting to the web evolution is considered. In this perspective, a *search engine web evolution rating mechanism* is exploited in order to reward efficient search services and penalize those that fail to perform effective information management.

The rest of the paper is organized as follows. Section 2 presents the fundamental concepts of our proposed third-party results ranking mechanism, aiming to offer an efficient way of personalized information resource search and acquisition. In Section 3 the search engine web evolution rating mechanism is described in a detailed manner in the context of capture-recapture experiments used in wildlife biological studies. Section 4 provides the formal description and mathematical formulation of the investigated third-party web results reliability rating system, while the browsing behaviour monitoring system is given in full detail. In Section 5 the overall ranking mechanism is mathematically formulated. Section 6 provides a set of results indicative of the performance of our proposed scheme. Finally, in Section 7 conclusions are drawn and future plans are given.

2 FUNDAMENTAL CONSIDERATIONS FOR THE PROPOSED RANKING MECHANISM

Assuming the presence of *M* Search Engine Services (SESs) each providing *N* web results (WRs) to the Meta-Search Engine Service (MSES) with respect to a specific user information resource request, MSES can combine and present to the user the web results acquired in re-organised manner on a basis of a third-party result ranking mechanism. The proposed mechanism rates the WRs in accordance with a weighted combination of the evaluation of the quality of each SES returned WR and an estimation which takes into account whether the user expectations concerning the specific WR raised in the past have been met. In our approach the first factor constitutes the *performance related factor,* while the second factor contributing to the overall WR ranking is referred to hereafter as the *reliability related factor*. The performance factor is introduced in order to take into account in our model the expected quality of each WR as given by each SES. To this end, the WRs ranking returned by each SES is considered. Considering the fact that the precision over the first recall level (top-ranked results) as given by each SES may be low, the reliability factor is considered in order to reflect whether WRs finally provide to the user the information resource that corresponds to his/her personalized requirements, preferences and constraints. The WR's reliability is reduced whenever the WR does not come up to the user expectations.

For the evaluation of the WRs reliability, a *collaborative WR reputation mechanism* is used, which helps estimating WRs quality and predicting the future (expected) usability, taking into account their past performance in consistently satisfying user expectations. In research literature, reputation mechanisms are employed to provide a "soft" security layer (considered to be sufficient for many applications [14]) by establishing trust relationships among system participants and/or choosing reliable resources [15], [16], [17], [18], [19], [20], [21], [22], [23]. Feedback received from participants related to an entity's past behaviour may be formulated as a reputation measure exploiting learning from experience concepts [24], [25], [26], [27]. The reputation related information obtained may be used by the parties in order to adjust their decisions and behaviour. In general, a reputation system is considered to sustain rational cooperation and serve as an incentive for good behaviour because good players are rewarded by the society, whereas bad players are penalized [14]. In this study, a reputation mechanism is exploited in order to collectively determine if a specific WR corresponds to user requirements and preferences with respect to a specific information resource search request. The reliability criterion is motivated by the fact that there may be different levels of user satisfaction with respect to the various WRs. In this respect, there may be WRs that, in principle, do not meet user requirements and preferences. Hence, recording the previous experience can easily assist the MSES in deciding how to present to the user the results obtained from the SESs. In our approach it has been assumed that past search behaviour is an indicator of the user's future behaviour, as a basis for user modelling.

In essence, for the formation of the WRs reliability ratings (and overall ratings) a centralized approach has been adopted (i.e., the MSES maintains and updates a collective record of the SES WRs reputation ratings, after taking into account each user view on the WRs performance). User's experience on the WRs performance is formed taking into account two factors. First, the time spent for his/her exploration as well as the 'depth' of the search. Time is considered as an important factor in determining user satisfaction upon a specific WR, since the more time the user spends exploring a specific result, the more this result is possible to be relevant and vice versa. As depth we have considered the number of hyperlinks used from the initiation of the search with starting point the SES WR result, until the session is closed. Every time the user browses a URL from the provided SES WRs, upon the end of the session, a reward function is calculated based on the aforementioned features, which is exploited in order to respectively update the WRs reliability value.

The proposed scheme is collaborative in the sense that it considers information acquired from various users in order to determine the reliability rating of each WR, enabling thus WRs reliability rating formation in a time efficient manner. At this point it should be noted that in the context of this study it is assumed that all users posing a specific search request for information resources have virtually the same information needs (in other word their profiles match). Additionally, the reliability value of each WR is formed irrespective of the SES that provided it. WRs reliability related information is acquired from each user session in a fully transparent way, without any interference in the user's browsing behaviour. Specifically, the user's personalized interaction patterns are monitored within the context of his/her sessions with a SES, while MSES results are presented to the user in the form of a text paragraph regarding the URL (as most web search engines do) without labelling their source, ensuring this way that the user is completely unbiased to a preference that may has in a particular SES.

The highly dynamic nature of the web necessitates effective information management. Thus, SESs should adequately adapt to the web evolution by indexing new information as quickly as possible and constantly checking the validity of their results (information resources do not live for ever or they are moved to another location or they are renamed). In general, SESs may demonstrate a different performance level with respect to the aforementioned issues. For example, results that no longer exist may be provided (dead links, errors 404), active but temporarily unavailable results may be given (web server internal errors, bad gateway, service/host unavailability), while new or updated information may be incorporated at a different pace by various SESs. This fact is taken into account in order to reward efficient SESs and penalize those that fail to perform effective information management. To this respect, a *SESs Web Evolution Rating Mechanism* is utilized in order to assign a ranking value to each SES, reflecting its ability to follow the dynamic nature of the web.

A learning period is required in order for the MSES to obtain fundamental information for the WRs. In case reliability specific information is not available to the MSES the reliability related factor is not considered for the WRs re-organisation. It should be noted that the reputation mechanism comes at the cost of keeping reputation related information at the MSES and updating it after each user session has taken place.

3 SEARCH ENGINE SERVICES WEB EVOLUTION RATING MECHANISM

In accordance with the aforementioned concepts and mechanisms, we take under consideration in our proposed scheme, the refreshness ability of third-party search services, or in other words, the ability of each SES to adapt to web evolution incidents by updating its indices and catalogues frequent enough. In this section, we describe in full detail the proposed SES rating mechanism that monitors how well the used search services adapt to the changes that occur on the web. This mechanism is put into the context of capture-recapture experiments used in wildlife biological studies, where animals are captured, marked and finally released on several trapping occasions. If a marked animal is captured on a subsequent trapping occasion, it is said to be recaptured. Based on the number of marked animals that are recaptured, using statistical models and their estimators, evolution sizes (such as the total population, well as the birth rate, the death rate and the survival rate of the animals under study) can be estimated. In this paper we use the Robust Design as an extension of the Jolly-Seber capture-recapture model [28], [29].

3.1 CAPTURE-RECAPTURE EXPERIMENTS: A DESCRIPTION

In a capture recapture experiment the sampling process is divided into i primary sampling periods, each consisting of j secondary sampling periods. Between primary periods the population is open to births and deaths, while on the other hand, between secondary sampling periods no births or deaths occur [30]. In a secondary sampling period a set of species is randomly selected according to a specific protocol (described in the following subsection), and accordingly marked as selected. The population is then mixed up again, and after an explicit time interval the next secondary period occurs, until the end of the last j^{th} secondary sampling period (Figure 1). Secondary samplings are close to time in order to assume that the sampled populations are closed. In other words, no losses or gains occur during these time intervals and each trapping occasion is also considered as closed. Conversely, time intervals between the primary sampling periods should be long enough so as evolution incidents can occur.

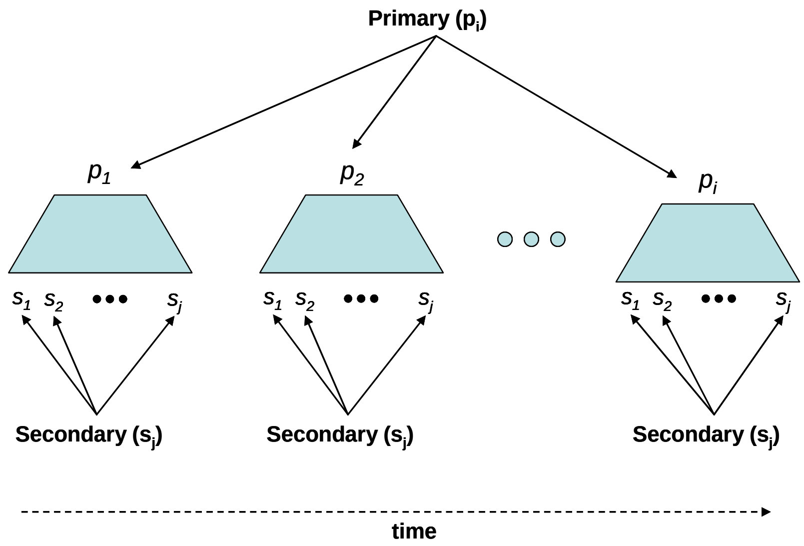

Figure 1. The basic structure of the followed capture recapture method (as described in [31])

3.2 Capture-recapture experiments on the web

In order to adopt the real-life experiments in our method, we considered the assumptions given in [32]. Yet, in our case, third-party WRs are the individuals under study, while the population consists of the set of the results provided by the MSES.

One basic consideration is that each third-party WR, which is present in the population (either marked or unmarked) during the time of the i^{th} sample, has the same probability p_i of being captured. This means that in the real life experiment the traps are set up for a specified amount of time, assuming that all animals have the same probability of being trapped. In our case, animal species correspond to different third-party WRs derived from different queries and SESs, while traps correspond to our sampling method. Thus, we had to ensure that third-party WRs of different queries have the same probability of being captured during the sampling procedure. Additionally, every marked third-party WR right after the i^{th} sample must have the same probability of survival φ_i until the next sampling instance. Assuming that marks are not lost and/or ignored, and are virtually instantaneous then Equations 1, 2 and 3 define the absolute value of births B_i, and the survival rate φ_i as well as the birth rate b_i of the tested population respectively. In these equations M_i is the number of marked third-party WRs in the population at the time where the i^{th} sample is collected, m_i corresponds to the number of the marked WRs captured in the i^{th} sample, N_i is the total number of third-party WRs in the population at the time where the i^{th} sample is collected, n_i is the total number of WRs captured in the i^{th} sample, B_i stands for the total number of new activated results entering the population between the i^{th} and $(i+1)^{th}$ samples and still remain in the population at the time $(i+1)^{th}$ sample is collected and finally R_i is the number of the n_i WRs that were released after the i^{th} sample.

$$\tilde{\phi}_i = \frac{\tilde{M}_{i+1}}{\tilde{M}_i - m_i + R_i} \tag{1}$$

$$\tilde{B}_i = \tilde{N}_{i+1} - \tilde{\phi}_i(\tilde{N}_i - n_i + R_i) \tag{2}$$

$$\tilde{b}_i = \frac{\tilde{B}i}{\tilde{N}_i} \tag{3}$$

We must note here that in our paradigm, survival rate (φ_i) corresponds to the portion of recaptured WRs between subsequent sampling periods, which were not updated (they still survive with the same unchanged content), while birth rate (b_i) corresponds to new WRs, which entered the tested population. In other words the bigger and lesser the birth and survival rates are the better we consider that the search service adapts to the web evolution.

According to the aforementioned concepts, the sampling protocol adopted is as follows. During the secondary sampling periods the mechanism records the users' submitted queries. The record fields are continuously updated, while its size works according to the leaky bucket model in the steady state mode, keeping the last S queries submitted by the users. We then parse the queries, and randomly select some of them under a probability value p_1. In the sequel, these queries (approximately $p_1 * S$) are submitted to the used web search services. The Top-*T* results (*ToT*) for each of the M web search services are collected, and the duplicate fields are removed. Then, we randomly select some of these results according to a second probability value p_2, thus allowing to each sampling instance to have the same probability of being included in each sample, independently of the instances that have already been sampled. This probability value is given by the product $p_1 \cdot p_2$, satisfying the first fundamental assumption, which requires that each third-party result in the population (either marked or unmarked) during the time of a sampling occasion i, must have the same probability of being captured ($p_i = p_1 \cdot p_2$, where $i = 1,2,...k$). After continuous secondary sampling periods we are able to estimate the survival and the birth rates.

Finally, in order to measure the ability of the tested web search services to adapt in the evolution that occur on the web, we defined the *WEAS* (Web Evolution Adaptation Score) as the product of the birth and survival rates between the subsequent primary sampling periods (Equation 4). This score is assigned for each SES after the end of the secondary sampling periods.

$$WEAS(b,\varphi) = b_i * \varphi_i^{-1} \tag{4}$$

4 WEB RESULTS RELIABILITY RATING SYSTEM FORMULATION

Concerning the formation of the web result WR_i ($i = 1,..., M \cdot N$) reliability rating $RR_{post}(WR_i)$, the MSES may rate WR_i after a user session d has taken place at time t_d in accordance with the following equation:

$$RR_{post}^{t_d}(WR_i) = RR_{pre}(WR_i) + k_r \cdot l(RR_{pre}(WR_i)) \cdot \{rr(WR_i) - E[rr(WR_i)]\} \tag{5}$$

where $RR_{post}(WR_i)$ and $RR_{pre}(WR_i)$ are the web result WR_i reliability based rating after and before the updating procedure. It has been assumed that $RR_{post}(WR_i)$ and $RR_{pre}(WR_i)$ lie within the $[0,1]$ range, where a value close to *0* indicates a web result that does not satisfy the user. $rr(WR_i)$ is a (reward) function reflecting the level of user satisfaction at the current session and $E[rr(WR_i)]$ is the mean (expected) value of the

$rr(WR_i)$ variable. In general, the larger the $rr(WR_i)$ value, the more satisfied is the user with the web result WR_i, and therefore the more positive the influence on the rating of the WR_i. Factor k_r ($k_r \in (0,1]$) determines the relative significance of the new outcome with respect to the old one. In essence, this value determines the memory of the system. Small k_r values mean that the memory of the system is large. However, greater usability will gradually improve the web results WR_i reliability ratings. $l(RR_{pre}(WR_i))$ is a function of the WR_i reputation rating $RR_{pre}(WR_i)$ and is introduced in order to keep the WR_i rating within the range $[0,1]$. In the current version of this study, $l(RR_{pre}(WR_i)) = \frac{1}{1-e} \cdot [1 - \exp(1 - RR_{pre}(WR_i)]$, for which it stands $\lim_{RR_{pre}(WR_i) \to 0} l(RR_{pre}(WR_i)) \to 1$ and $\lim_{RR_{pre}(WR_i) \to 1} l(RR_{pre}(WR_i) \to 0$.

It should be noted that web results deterioration of its previous quality leads to a decreased post rating value, since the $\{rr(WR_i) - E[rr(WR_i)]\}$ quantity is negative. The $rr(WR_i)$ function may be implemented in several ways. In the context of this study, it was assumed without loss of generality that the $rr(WR_i)$ values vary from 0.1 to 1, while it is calculated on the basis of two factors: time spent exploring a specific web result WR_i and the respective depth of the search. The formulation of the reward function $rr(WR_i)$ is given in the following subsection.

4.1 REWARD FUNCTION FORMULATION (BROWSING BEHAVIOUR MONITORING MECHANISM)

The proposed personalization algorithm weights the relevancy of the provided web results, based on the users' web search interactions. As mentioned previously, we have assumed that past search behaviour is an indicator of the user's future behaviour. The construction of the personalized browsing behaviour is performed in a totally transparent way, while the merged WRs are presented without labelling their source. The only feedback the user receives is a text paragraph regarding the URL, as most of web search engines do. Personalization patterns are recorded and updated continuously according to the WRs visited by the user, the time spent for their exploration as well as the depth link of the investigated results. Thus, the user's profile is also adjusted to any possible changes in respect to his/her navigation patterns.

In order to measure the similarity of a WR in respect to the user's behaviour, we use two probabilistic functions. These functions assign a probability value according to the time the user spends for information exploration as well as according to the depth d of the investigated WR, where d denotes the distance in hops (subsequent hyperlinks) between the investigated third-party WR and the link reached when the session is closed. In our approach, the time needed for exploring a third-party result is quite important, since the more time the user spends for exploring a specific result, the more this result is possible to be relevant and vice versa. Thus, the time between following visits during a search session, was modelled according to a lognormal distribution. This kind of distribution was selected among others, given that this distribution fits with the results made in respect to a large user browsing behaviour analysis as described in [33]. As far as the time spent for information exploration is concerned, we relied on the same survey [33], where the authors experimentally concluded that, on average, users made 2.4 searches (hops) per session. In addition, the average search session duration was approximately 2 minutes (1 minutes and 50 seconds). In a similar study described in [34], the authors noticed that nearly 72% of the users spend five minutes at the most during a search session. Taking into account the fact that not all users are familiar with a search engine interface, thus, they may spend more time in comparison with other users, as well as, in order to avoid misjudges in the scoring of the result due to idle activity periods in the user's work, we selected to monitor the user's browsing behaviour during the first five minutes he/she accesses the web resource. Based on the above, the browsing behaviour score (*BBS*) is provided according to Equation 6.

$$BBS_{WR_i}(t,d) = rr(WR_i) = \frac{\max[P_t \cdot P_d]}{\sum_{d=1}^{l} \int_{t=0}^{T} (P_t \cdot P_d)\,dt} \tag{6}$$

where $P_d(x) = \frac{e^{-((\ln((x-\theta/m))^2/(2\sigma^2))}}{(x-\theta)\sigma\sqrt{2\pi}}$ stands for the lognormal distribution probability density function, where $\theta = 0.5$, $m = \sigma = 1$, while $P_t(x) = \Phi(\frac{\ln(x)}{\sigma})$ is the lognormal cumulative distribution function, where Φ is the cumulative distribution function of the normal distribution and θ, m, σ are the location, scale and shape parameters respectively.

5 WEB RESULTS RANKING MECHANISM FORMULATION

The target web result WR_i ($i = 1,...,M \cdot N$) provided by search engine service SE_j ($j = 1,...,M$) may be rated by the evaluator MSES at time t_c that a user request has to be served in accordance with the following formula (Equation 7):

$$WPR_{SE_j}^{t_c}(WR_i) = NWEAS_{SE_j} \cdot \left\{ w_p \cdot PR_{SE_j}^{t_c}(WR_i) + w_r \cdot RR^{t_c}(WR_i) \right\} \tag{7}$$

where $WPR_{SE_j}^{t_c}(WR_i)$ denotes the overall rating of the web result WR_i provided by SES SE_j at time instance t_c.

As may observed from Equation 7, the rating of the target WR_i is a weighted combination of two factors. The first factor contributing to the overall WR_i rating value (i.e., $PR_{SE_j}^{t_c}(WR_i)$) is based on the performance of the WR_i as given by SES SE_j and forms the performance related factor. In a similar manner to the WRs reliability rating value, it has been assumed that $PR_{SE_j}^{t_c}(WR_i)$ lie within the [0,1] range, where a value close to *0* indicates that the performance of the web result as given by the SES is low. In the context of this study, $PR_{SE_j}^{t_c}(WR_i)$ is given by the following expression (Equation 8):

$$PR_{SE_j}^{t_c}(WR_i) = 1 - \frac{k}{N} \tag{8}$$

where k is the rank level of WR_i as returned to the MSES by SES SE_j at time instance t_c and N is the number of WRs provided by SE_j.

The second factor (i.e., $RR^{t_c}(WR_i)$) depends on user satisfaction stemming from WR_i, collectively formed considering all user service search requests in the past. In essence, this factor constitutes the reliability related factor and is a function of the $RR^{t_d}_{post}$ as given by Equation 5, where t_d denotes the time instance at which WR_i was last time accessed and the respective reliability value was accordingly updated. A wide range of functions may be defined. We restrict our attention to the polynomial family of functions. Other functions could be defined as well. Equation 9 provides a formal model of the polynomial related family of functions concerning the $RR^{t_c}(WR_i)$ reliability rating.

$$RR^{t_c}(WR_i) = cr(RR^{t_d}_{post}(WR_i)) \cdot [1-(\frac{t_c - t_d}{t_c})^{1/\vartheta}] \cdot RR^{t_d}_{post}(WR_i) \tag{9}$$

where $cr(RR^{t_d}_{post}(WR_i)$ is the credibility of the WR_i reliability rating $RR^{t_d}_{post}(WR_i)$ given by the following Equation 10.

$$cr(RR^{t_d}_{post}(WR_i) = \frac{N_U(WR_i) \cdot \sum_{u=1}^{N_U} N^u_H(WR_i)}{N_U \cdot \sum_{u=1}^{N_U} N^u_H} \tag{10}$$

where $\sum_{u=1}^{N_U} N^u_H(WR_i)$ is the number of hits WR_i receives considering all user sessions, $\sum_{u=1}^{N_U} N^u_H$ is the total number of hits considering all WRs and all user sessions, $N_U(WR_i)$ is the number of users that begun a session with WR_i, and N_U is the total number of users using the MSES till time instance t_d. It should be noted that, for one user u, $N^u_H(WR_i)$ is assumed to be a boolean variable (i.e., $N^u_H(WR_i) \in \{0,1\}$). More than one hits for the same WR originating from the same user within the context of a specific MSES request are not taken into account for the calculation of the credibility value of the reliability rating of a specific WR (that is a user may access or not a specific web result). For the WR reliability evaluation we have used the user session achieving the higher reward function (in accordance with Equation 6).

As may observed from Equation 9, it stands $\underset{t_c \to t_d}{RR^{t_c}(WR_i)} \to cr(RR_{post}^{t_d}(WR_i)) \cdot RR_{post}^{t_d}(WR_i)$ and $\underset{t_c >> t_d}{RR^{t_c}(WR_i)} \to 0$. Specifically, the bigger the quantity $t_c - t_d$ is, the lower is the reliability value considered for the WR_i. Equation 9 in essence models the fact that more recent user interactions with a specific WR should weigh more in the overall WR ranking evaluation. As it may be observed from Equation 9, these families of functions represent an infinite number of different members, one for each value of ϑ. Parameter ϑ has been included in order to highlight the different patterns with respect to the adopted rate of decrease. For example, adopting a Boulware policy [35] could lead to minor modification (decrease) of the reliability rating, until $\frac{t_c - t_d}{t_c} \to 1$ (i.e., $\frac{t_d}{t_c} \to 0$), whenupon, the minimum reliability value is assumed. Otherwise, exploiting the Conceder policy [36] could lead to the minimum reputation value in quite a short time period (the quantity $t_c - t_d$ is quite small).

Weight $NWEAS_{SE_j}$ is the SES web evolution rating value as given by Equation 4 normalized so as to lie within [0,1] range and is introduced in order to reward SES that perform efficient information management, while penalizing those that fail to follow the web evolution dynamics. Finally, weights w_p and w_r provide the relative value of the anticipated WR performance as given by each SES and the reliability related part. It is assumed that weights w_p and w_r are normalized to add up to 1 (i.e., $w_p + w_r = 1$). It should be noted that in certain variants of the problem one of the two factors may be ignored. In the following sub-section an illustrative example of the WR re-organization process, so as to become comprehensive to the reader.

5.1 *RE-ORGANIZING THE WEB RESULTS: AN ILLUSTRATIVE EXAMPLE*

During each secondary sampling period, a set of about $S \cdot p_1 \cdot p_2 \cdot ToT$ web results is examined for each of the search engine services considered. On the basis of equation (7), the overall rating of each returned result is estimated, taking into account both the performance related factor (equation (8)) and the reliability related factor (equations (5)-(6)-(9)-(10)), while the web evolution adaptation score of each search engine service is considered (equation (4)). Assuming that the MSES possesses an accurate picture concerning the web results reliability factor (that is after the learning period), a single list of merged-results is formed, after combining the returned results, removing duplicate web results (in the current version of this study, the ones provided by the most efficient search engine service with respect to information management are kept) and re-organizing the web results so as the ones with the higher estimated overall rating occupy the first ranking positions in the merged-result list.

Following, an illustrative example is provided, so as the proposed MSES mechanism becomes comprehensive to the reader. The individual result ranking positions for the three SESs considered (namely Google, MSN, and Yahoo!) are as presented in Table 1 (1^{st} top-ranked result of Google is G_1, 2^{nd} is G_2, 1^{st} top-ranked result of MSN is M_1, 2^{nd} is M_2, 1^{st} top-ranked result of Yahoo! is Y_1, 2^{nd} is Y_2, etc.). After removing duplicate web results, on the basis of the acquired WRs overall rating, the MSES returns to the user a combined web result list with G_3 occupying the first ranking position, followed by M_2 in the second ranking position, G_1 in the third ranking position, etc.

Index	**Search Engine Service**			**MSES Combined WR List**
	Google	**MSN**	**Yahoo!**	
1	G_1	M_1	Y_1	$\boldsymbol{G_3}$
2	G_2	M_2	Y_2	$\boldsymbol{M_2}$
3	G_3	M_3	Y_3	$\boldsymbol{G_1}$
4	G_4	M_4	Y_4	$\boldsymbol{Y_1}$
5	G_5	M_5	Y_5	$\boldsymbol{G_2}$
6	G_6	M_6	Y_6	$\boldsymbol{M_1}$
7	G_7	M_7	Y_7	$\boldsymbol{Y_2}$

8	G_8	M_8	Y_8	$\boldsymbol{Y_3}$
…	…	…	…	**…**

Table 1. MSES Mechanism: Web Result Combined List Formation

6 PERFORMANCE EVALUATION – EXPERIMENTAL RESULTS

In order to evaluate our proposed mechanisms, a set of 47 queries (parameter S in the sampling protocol description - Section 3.2) regarding dermatology was created by a user group consisting of four medical researchers, who had their research focused in this field. We wanted to quantify the average precision for the top-50 (parameter ToT in the sampling protocol description - Section 3.2) merged third-party results given by $M = 3$ SESs (namely, Google, MSN and Yahoo), over different recall levels using the predefined 47 queries. The experiments were conducted according to sampling protocol described in section 3.2, where the time interval between two following primary sampling periods was set equal to 28 days, while the respective time for subsequent secondary sampling periods was two days. Time interval between primary sampling periods was set to 28 days due to the fact that the capture recapture methodology can easily follow and adapt to the changes that occur on the web, if the difference between subsequent sampling periods is above 26 days [37]. Thus, in order to reduce possible errors, we extended this time-window for two additional days (28 days). Four secondary sampling periods were used in the context of the experiments performed. In essence, the secondary periods took place at the beginning of the first, the third, the fifth and the seventh day after the initiation of each primary one. In other words, the time needed for the completion of a primary sampling period was nearly six days. The probability values p_1 and p_2 were set equal to 0.3. Thus, during a secondary period we selected nearly 14 queries ($S \cdot p_1$=47 queries x 0.3), which in the sequel were submitted to the three third-party web search services. A subset of the $ToT = 50$ returned results were randomly selected with probability p_2. Overall, we expected to examine nearly 210 results (47 queries x 0.3 x 50 results x 0.3) for each search engine service considered during the secondary sampling periods. Probability values p_1 and p_2 are actually fine-tuning parameters, and are mainly used in order to generate the population under study.

Our aim is to provide a measure over the overall performance of the proposed MSES for the three search engine services used on a six-day basis (the time interval between subsequent primary sampling periods). During the learning period, the MSES, having removed duplicate web results, provides users with a combined list, without however re-organizing their respective positions. In essence, the 1st top-ranked result of SES 1 (i.e., Google) occupies the first ranking position of the merged list, 1st top-ranked result of SES 2 (i.e., MSN) occupies the second ranking position, 1st top-ranked result of SES 3 (i.e., Yahoo!) occupies the third ranking position, 2nd top-ranked result of SES 1 (Google) occupies the fourth ranking position, etc.. Based on users' personalized browsing behaviour monitoring mechanism (equation (6)), the WRs reliability ranking is estimated in accordance with equation (5). Additionally, the web evolution adaptation score for each SES is formed on the basis of equation (4). Having formed an accurate picture concerning the web results reliability factor (that is after the learning period), the MSES forms a single list of merged-results, re-organizing the web results so as the ones with the higher estimated overall rating on the basis of equation (7) occupy the first ranking positions in the merged-result list (as illustrated in the previous section). At this point it should be noted that the construction of the users' personalized browsing behaviour as well as the formation of the search engine services' web evolution adaptation score is achieved in a fully transparent manner, as the whole procedure was running in the background, without requiring users' interference. For the estimation of the time the users spent for exploring the obtained results, we used a software module (client-side agent) that reproduces the Unix timestamp (the amount of seconds since January 1 1970 00:00:00 GMT) for each instance when the web resource is browsed. Thus, for each visited web result, we calculated the difference between subsequent recorded timestamps for each hop during a search session.

Our proposed WR ranking mechanism performance assessment was made with Precision-Recall (PR) diagrams, having averaged the respective PR values in a monthly basis from June to November of 2007. We observed that the precision over different recall levels for the top-fifty returned results was increased for all four users, who submitted the same 47 queries and examined the returned results over a six-month period (June 2007 to November 2007). Figures 2 up to 5 illustrate the positive influence of the collaborative third party web result rating system for each one of the four users independently, while Figure 6 depicts the averaged PR values for all users and during the same tested period. After the end of the assessment and after analyzing the depicted results, we observed a considerable increase in the measured precision values over different recall levels (0, 0.1, …, 1.0).

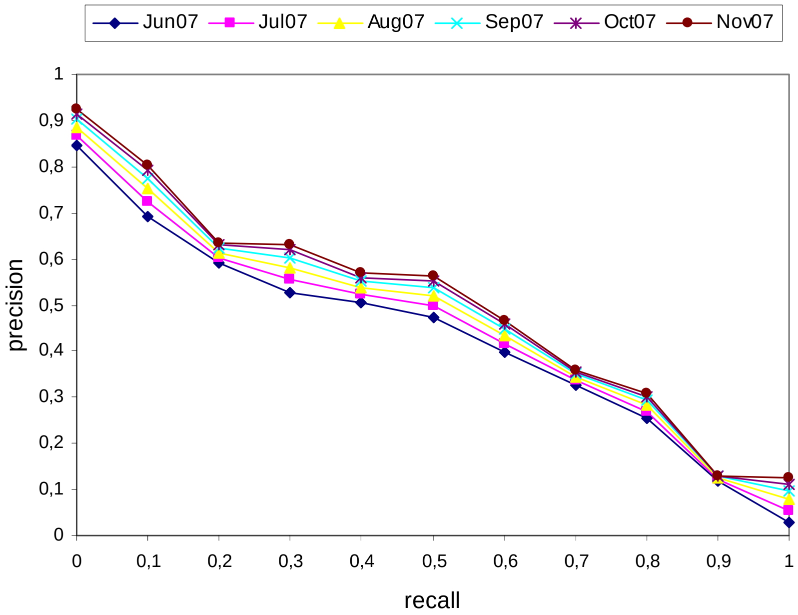

Figure 2. Precision-Recall level for User #1 [Jun 07 – Nov07]

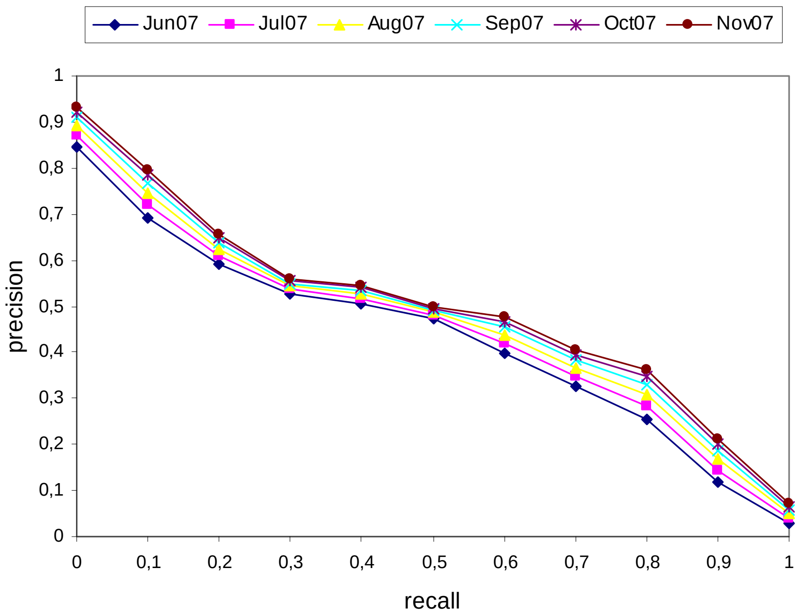

Figure 3. Precision-Recall level for User #2 [Jun 07 – Nov07]

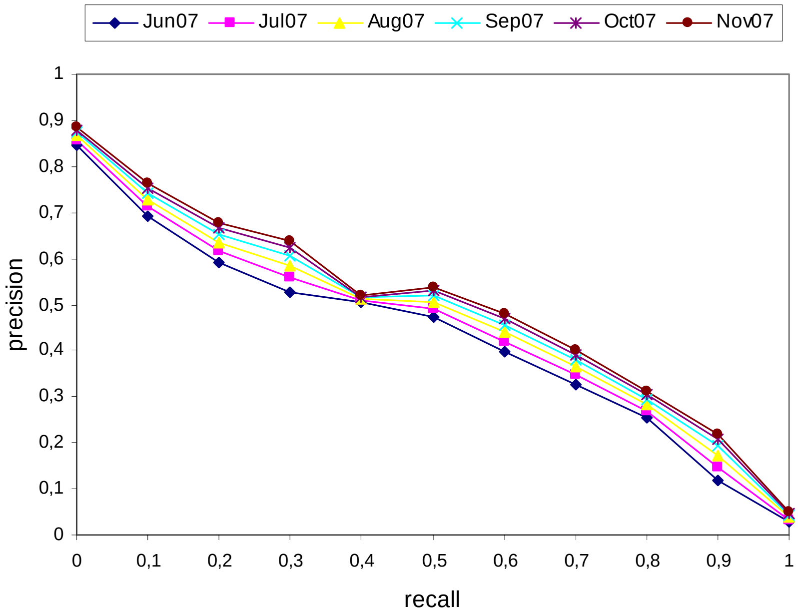

Figure 4. Precision-Recall level for User #3 [Jun 07 – Nov07]

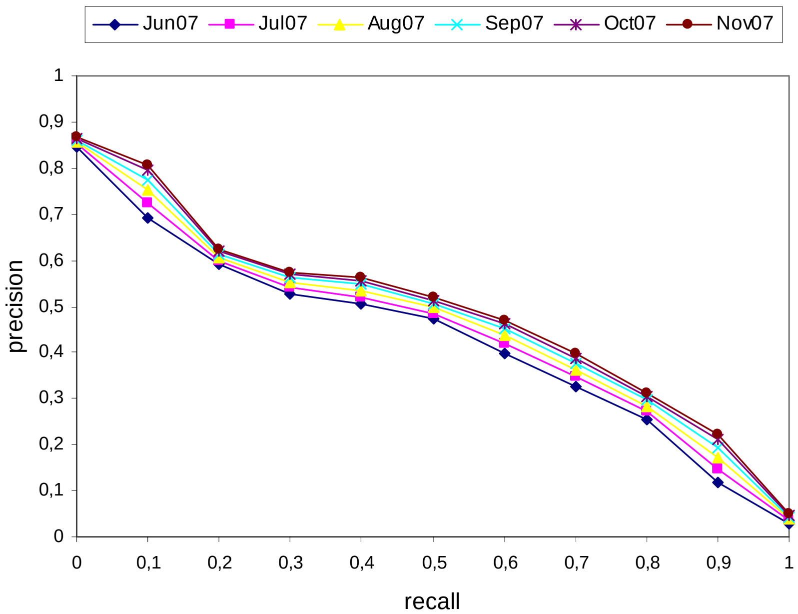

Figure 5. Precision-Recall level for User #4 [Jun 07 – Nov07]

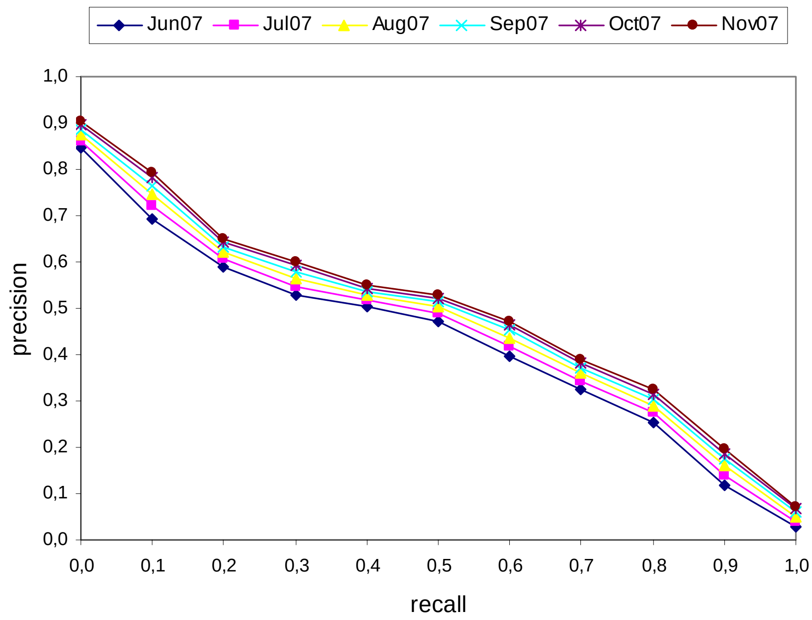

Figure 6. Averaged Precision-Recall values [Jun 07 – Nov07]

Table 2 depicts the improvement (in percentage) for all averaged PR values in a monthly scale from June up to November of 2007. According to this table, we have an average improvement at the levels of 4.44% during the first month (from June to July 2007). In the sequel, an additional improvement of 3.64% for the averaged PR values occurs for the time interval between August and July of 2007, resulting in an improvement of more than 8% (8.24%) for the two month-period between August and June of 2007. Finally, the total improvement in the averaged precision values from the initial PR values and the final averaged values for all users for the six-month period between June and November of 2007 reached the levels of 16% (15.85%), indicating that our collaborative WR reliability rating system formulation works perfectly in joint use with the browsing behaviour monitoring mechanism and the search engine services web evolution rating mechanism, which monitor the user's browsing behaviour as well as the ability of the third-party search services to follow the web evolution dynamics. It was also noticed that the improvement in percentage values between subsequent months (June-July, July-August, August-September, September-October and October-November), was decreased over the time. This was a desired effect of our proposed scheme, since it proves that our MSES WR rating system continuously adapts to the users' preferences and browsing behaviour.

% average improvement of Precision over Recall						
	Jun07	Jul07	Aug07	Sep07	Oct07	Nov07
Jun07		4,44	8,24	11,42	13,95	15,85
Jul07	-		3,64	6,68	9,11	10,93
Aug07	-	-		2,93	5,27	7,03
Sep07	-	-	-		2,28	3,98
Oct07	-	-	-	-		1,67
Nov07	-	-	-	-	-	

Table 2. Precision improvement between different time intervals (in months)

7 CONCLUSIONS

In accordance with efficient web search service operation objectives, the aim of this paper is to propose enhancements to the sophistication of the functionality that can be offered by search engine services. Specifically, a meta-search third-party web result ranking mechanism is proposed, which enables for personalized information acquisition, taking into account the user's preferences, requirements and constraints, implicitly, by monitoring his/her navigation behaviour. The proposed mechanism is capable of adapting over the continuous changes that occur on the web, rewarding search engines performing effective information management, while penalizing those that fail to follow the dynamic nature of the web. Transparency is achieved for both personalization and web evolution adaptation mechanisms, requiring virtually none effort from the user's part.

In essence, the proposed meta search engine rates, re-organises and combines the results acquired from search services for a specific user information resource request in accordance with a weighted combination of a performance related factor (tightly related to the web result ranking as given by the search engine service) and a reliability related factor (corresponding to the user satisfaction stemming from the specific web result that he/she browses), while the performance of each search engine with respect to adequately adapting to the web evolution is taken into account. For the evaluation of the web results reliability, a collaborative reputation mechanism is utilized, which helps estimating their quality and predicting their future usability, taking into account their past performance in consistently satisfying user expectations. Experimental results have been obtained over a six month period time (June to November 2007) by using a set of 47 queries regarding dermatology created by a group of four medical researchers focused in this field. It was observed that the precision over several recall levels was increased, for all time intervals between subsequent months, while our proposed third party web result ranking mechanism, resulted in an average improvement of approximately 16% for the whole time period.

Directions for future work include, but are not limited to the following. First, the realization of further wide scale experiments considering user groups with different information needs, so as to evaluate the applicability and the response of the framework presented herewith and second, the server-side implementation of our proposed mechanism. Finally, we consider the implementation of a collaborative web result ranking mechanism with users not necessarily having matching profiles concerning their information resources preferences and requirements.

REFERENCES

[1] Pitkow J., Schutze H., Cass T., Cooley R., Turnbull D., Edmonds A., Adar E., and Breuel T., Personalized search. Communications of the ACM, 45(9):50–55, 2002.

[2] Liu F., Yu C., and Meng W., Personalized web search for improving retrieval effectiveness. IEEE Transactions on Knowledge and Data Engineering (TKDE), 16(1):28-40, 2004.

[3] Fitzpatrick L. and Dent M., Automatic feedback using past queries: social searching? In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 306–313. ACM Press, 1997.

[4] Google personalized search, http://labs.google.com/personalized

[5] My yahoo!, http://my.yahoo.com/?myhome

[6] Carroll J., and Rosson M., The paradox of the active user. Interfacing Thought: Cognitive Aspects of Human-Computer Interaction, 1987.

[7] Velayathan G., and Yamada S. (2006), Behavior-based web page evaluation. In Proceedings of the 15th International Conference on World Wide Web (WWW '06), pp. 841-842. ACM Press, 2006.

[8] Sugiyama K., Hatano K., and Yoshikawa M., Adaptive web search based on user profile constructed without any effort from users. In Proceedings of the 13th International Conference on World Wide Web, pp. 675–684, ACM Press, 2004.

[9] Page L., Brin S., Motwani R., and Winograd T., The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[10] Haveliwala T. H., Topic-sensitive pagerank. In Proceedings of the 11th International Conference on World Wide Web, pp. 517–526. ACM Press, 2002.

[11] Gordon M., Weiguo F., Pathak, P., Adaptive Web Search: Evolving a Program That Finds Information, IEEE Intelligent Systems, [see also IEEE Intelligent Systems and Their Applications], 21(5):72 – 77, 2006.

[12] Spinka A., Jansenb B. J., Blakelyc C., and Koshmand S., A study of results overlap and uniqueness among major Web search engines, Information Processing & Management, 42(5):1379-1391, 2006.

[13] Bar-Ilana J., Mat-Hassanb M., and Leveneb M., Methods for comparing rankings of search engine results, Computer Networks, 50(10):1448-1463, 2006.

[14] Zacharia G., and P. Maes, Trust management through reputation mechanism. Applied Artificial Intelligence Journal, 14(9): 881-908, 2000.

[15] Louta M., Roussaki I., and Pechlivanos L., Reputation Based Intelligent Agent Negotiation Frameworks in the E-Marketplace. In Proceedings of International Conference on E-Business (ICE-B 2006), pp. 5-12, INSTICC press, 2006.

[16] Damiani E., Vimercati S., Paraboschi S., Samarati P., and Violante F., A Reputation-Based Approach for Choosing Reliable Resources in Peer-to-Peer Networks. In Proceedings of 9th ACM Conference on Computer and Communications Security, pp. 207-216. ACM Press, 2002,.

[17] Huynh T., Jennings N., and Shadbolt N., Certified Reputation: How an Agent can Trust a Stranger. In Proceedings of the 5th International joint conference on Autonomous Agents and Multi-Agent Systems, pp. 1217-1224. Hakodate, Hokkaido, Japan, 2006.

[18] Mui L., Halberstadt A., and M. Mohtashemi, Evaluating Reputation in Multi-Agent Systems. Trust, Reputation, and Security: Theories and Practice (AAMAS 2002),

LNAI 2631 (R. Falcone et al;, Eds.), pp. 123-137. Springer-Verlag, Berlin Heidelberg, 2003.

[19] Sabater J., and C. Sierra, REGRET: Reputation in gregarious societies. In Proceedings of the 5th International Conference on Autonomous Agents, pp. 194-195. Montreal, Quebec, Canada, 2001.

[20] Xiong L., and Liu L., Reputation and Trust, Advances in Security and Payment Methods for Mobile Commerce, pp. 19-35. Idea Group Inc., 2005.

[21] Yu B., and Singh M., A social mechanism of reputation management in electronic communities, Cooperative Information Agents IV - The Future of Information Agents in Cyberspace (the 4th International Workshop on Cooperative Information Agents), LNCS 1860 (M. Klusch and L. Kerschberg, Eds.), pp. 154-165. Springer-Verlag Berlin Heidelberg, 2000.

[22] Yu B., and Singh M., Detecting Deception in Reputation Management. In Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 73-80. Melbourne, Australia, 2003.

[23] Zacharia G., Moukas A., and Pattie Maes, Collaborative Reputation Mechanisms in Electronic Marketplaces. In Proceedings of the 32nd Hawaii International Conference on System Sciences. Los Alamitos, CA, USA, 1999.

[24] Sutton R., and Barto A., Reinforcement learning: An introduction (Adaptive computation and machine learning). MIT Press, 1998.

[25] Michalski R., Carbonell J., and Mitchell T., Machine Learning: An Artificial Intelligence Approach. Morgan Kaufmann Publishers, 1986.

[26] Cherkassky V., and Mulier F., Learning from Data : Concepts, Theory, and Methods (Adaptive and Learning Systems for Signal Processing, Communications and Control Series). John Wiley & Sons, March 1998.

[27] Kaelbling L., Littman M., and Moore A., Reinforcement Learning: A Survey, Journal of Artificial Intelligence Research, 4: 237-285, 1999.

[28] Kendall W., Nichols J., and Hines J., Estimating temporary emigration using capture recapture data with Pollock's robust design. Ecology, 78: 563-578, 1997.

[29] Jolly, G., Explicit estimates from capture-recapture data with both death and immigration stochastic model, Biometrika, 52: 225-247, 1965.

[30] Pollock K., Nichols J., Brownie C., and Hines J., Statistical inference for capture-recapture experiments. Wildlife Monographs 107, 1990.

[31] Program MARK, A Gentle Introduction, 5th Edition, downloaded from http://www.phidot.org/software/mark/docs/book/

[32] Eguchi T, Statistic class notes (Topics in Ecological Statistics), Montana State University, based on book by Seber (1982; The estimation of animal abundance and related parameters. Second edition, Macmillan Publishing Co., New York, NY), http://www.esg.montana.edu/eguchi/pdfFiles/markRecapSummary.pdf

[33] Web Site Usability Metrics - Search Behavior and Search Trends, White Paper by Mondosoft Development Team, last updated May 2004, downloaded from http//www.mondosoft.com/SearchBehaviorWP.pdf

[34] Jansen B.J., Spink A., Koshman S., Web Searcher Interaction with the Dogpile.com Metasearch Engine, Journal of the American Society for Information Science and Technology, 58(5):744–755, 2007.

[35] Raiffa H., The Art and Science of Negotiation. Harvard University Press, Cambridge, USA, 1982.

[36] Pruitt D., Negotiation Behavior. Academic Press Inc., 1981.

[37] Anagnostopoulos I., and Anagnostopoulos C.N., Cache adaptation of Internet services and user modeling for improving ranking precision in third-party results, Second International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), pp. 122-127, London, UK, 2007.