Trust-based Testbed for P2P digital library

Rami S. Alkhawaldeh*
Department of Computer Information Systems
The University of Jordan
Aqaba, 77110
Jordan
r.alkhawaldeh@ju.edu.jo

Moatsum Alawida
School of Computer Sciences
University Sains Malaysia
11800 USM, Pulau Pinang
Malaysia
matsm88@yahoo.com

Issam AlHadid
Department of Business Information Technology
The University of Jordan
Aqaba, 77110
Jordan
issamfe@hotmail.com

Abstract. Retrieving relevant information from trust-based Peer-to-Peer (P2P) networks is a challenging research task. The users on P2P networks endeavor to ensure that the peers have to provide guaranteed relevant information (or documents) for their information needs. Recent researches in evaluating the trustworthiness or even the reputation of peers lack a well-formulated testbed. As such, building a robust testbed for evaluating trustworthy P2P networks motivates to propose a ground truth testbed as a contribution in P2P systems. In this paper, a trust-based testbed is developed especially in P2P digital libraries using TREC WT10g collection which has been used for evaluation in many P2P networks. The testbed contains a set of peers of trustworthy values for each peer that is estimated using peer rank approach. The statistical factors such as the distribution of peer trustworthiness, distribution of relevant documents and the location of relevant documents within each setting, and the importance of proposed testbed to reach 100% Recall in information retrieval are analyzed. The results are compared with two well-known testbeds.
Keywords: P2P, trustworthiness, page rank, peer rank, digital library.

1. Introduction

Peer-to-Peer (P2P) overlay network is a distributed computing system emerging as a popular way to share a huge amount of data [14]. The nature of the P2P

*. Corresponding author

overlay network gives the flexibility to form many application scenarios from open file sharing (content exchange) and resource sharing (CPU, storage and bandwidth) to digital libraries [8]. Digital Library (DL) application is considered one of the common applications of such a network for content storage and retrieval [16]. These libraries have common characteristics such as (i) contain a large number of documents, (ii) their documents are fairly uniformly distributed and (iii) they share the same popularity in the pattern of documents access. In order to build a testbed to evaluate the performance of IR systems related to DL applications, many factors should be taken into account such as the digital library trustworthiness in order to retrieve pure and high-quality materials. Although the traditional search engines that are centralized or distributed over self-organized P2P overlay networks mainly focus on improving the retrieval effectiveness and efficiency, the users of P2P networks concern with relevant and trustworthy documents. These documents are relevant to the issued query and at the same time are not corrupted or contain viruses to satisfy their information needs. Because of the content distribution in P2P information Retrieval (P2PIR) is subject to constant change as results of operations such as insertion, deletion, and/or modification, the quality of the published contents (i.e. their trustworthiness) becomes an important issue in these applications. This means that there are some peers more trustworthy than the others. Thus, trust-aware P2P overlay networks are concerned with the retrieval of not only relevant but also high-quality documents from trusted peers.

Several types of research have been dedicated to essentially improve and focus on the retrieval effectiveness of DL applications. In [10] proposed a hybrid P2PIR approach using a large scale testbed and a set of queries. The retrieval effectiveness is estimated where the efficiency of the system is evaluated based on the number of routed queries. They considered, in their testbed, the combined documents that have similar content (topic) by using the similarity-based soft-clustering algorithm. They only created the testbed depending on the topic of documents ignoring any factors that are related to documents or peers trustworthiness. In [7] proposed a cluster-based P2P architecture that is evaluated using six testbeds developed from WT10g collection as real suite large-scale testbeds [9]. In [9] considered three factors for building their testbeds: (i) document distribution across peer-collection (which follows a power-law distribution), (ii) the location of relevant documents (Recall) and (iii) coverage of the topics (precision). However, they did not mention or even tackled the trust factor associated with DL applications. The same testbeds are used by [1,2,3] for evaluating a set of routing techniques on a cluster-based architecture P2P network to retrieved relevant documents of a given query maintained into clusters (or super-peers). To our knowledge, there is no common ground truth testbed that is realistic to be used for P2P system evaluation depending on the trust factor. Therefore, an approach is proposed to build a testbed by measuring the importance of the peers (i.e. their trustworthiness) by estimating and assigning a trust value, named peer rank. These values facilitate routing the given query to the rele-

vant and trustworthy peers in providing high-quality contents. *The contribution here is suggesting a proposed testbed for P2P overlay architectures that focusing on peer trustworthiness evaluation. In particular, the testbed utilizes the well-known page rank algorithm to estimate the peers rank based on the connected graph between each other.*

The remainder of the paper is organized as follows. Section 2 presents research related to trustworthiness; Section 3 discusses the proposed approach on building the testbed. Section 4 presents the experimental settings and the data set used for constructing the proposed testbed and also analyzes the created testbeds from trustworthiness and effectiveness point of views, followed by Section 5 that concludes the findings of the paper and the future work.

2. Related works

Identifying the quality of documents stored in P2P network nodes has revealed a new concept in P2P information retrieval research area called entity trustworthiness. The trustworthiness of an entity could be defined as the quality of relevant documents or service providers for satisfying the user's information needs [3]. There are many techniques to estimate the trust values of peers, one of them is gathering the number of positive or negative feedback from the other peers to be exploited for estimating the reputation of a specific peer [17]. The trustworthiness of peers in P2PIR must be taken into consideration by researchers due to apparently the widespread applications that depend on P2P systems such as block-chain systems which might be a security mechanism for authentication and authorization [13]. In addition, the users in P2P networks have full permission to publish their content which might lead to overwhelming the network with malicious or selfish peers or even untrustworthy documents or feedback. Therefore, the necessity for robust techniques in order to filter such malicious (or selfish) peers or untrustworthy documents (or feedback) is high.

Many techniques have been proposed to compute the trustworthiness of documents and peers. For instances, EigenTrust [6] is a reputation-based system that depends on filtering out inauthentic files in P2P file-sharing networks. In this technique, a unique global trust value for each peer in the system is estimated based on the local opinions of all the other peers. The local trust value is a summation of user ranks that are gathered directly or indirectly of either positive or negative rate. In [15] proposed a multi-agent trustworthiness model where each agent maintains models of its acquaintances that describe their expertise (the quality of the services they provide) and sociability (the quality of the referrals they provide). These two works used their own testbeds to evaluate their systems and compute the trust factor through the lifetime of the system. In [12] suggested two trustworthiness methods to calculate the trustworthiness of a specific peer: subjective and objective trustworthiness. The subjective trustworthiness is estimated through direct interaction with the acquaintance peer where the objective trustworthiness is estimated by collecting trustworthi-

ness information from other peers. This is similar to the reputation-based trust model that uses experience or the experiences of others as a recommendation. In [18] also proposed to estimate documents trustworthiness depend on document reputation (objective trustworthiness) and peer trustworthiness (subjective and objective trustworthiness). In [4] built an approach that depends on the combination of some metrics to calculate the trust values of peers. The system suggested many factors to calculate the trustworthy peer value, where the increase in factors will force these systems to built by a small number of peers for evaluation. VectorTrust [20] is a trust management system for aggregating distributed trust scores, which is built on a Trust Overlay Network (TON) on the top of decentralized P2P networks. VectorTrust uses a Bellman-Ford based distributed algorithm for fast trust score aggregation.

Although there are many models for P2P overlay networks that cover the trust factors in these type of the systems, these proposed systems use their own testbeds for trust-based evaluation which are small in their nature. Therefore, they do not reflect the real environments and provide a useful testbed for comparison. As such, this motivates us to build a testbed simulate the trust values of peers. The proposed testbed is built based formally on the peers rank values; which is the number of access to a specific peer by the link structure of documents inside the peer as a factor for trustworthiness.

3. Trust-based Testbed for digital libraries

The proposed trust-aware testbed has the characteristics of real DL applications. The documents are distributed evenly over the peers with almost the same number of documents. This paper proposes a method to build a testbed that satisfies such characteristic called Digital Library Peer Rank (DLPR). The UWOR (Uniform WithOut Replication) testbed is used built by [9] to be used for systems with uniformly documents distribution, as initial testbed to build DLPR. The UWOR testbed contains 11,680 domains (or peers) that guarantee the equal number of documents on each peer for the proposed testbed. The process of building DLPR testbed has three phases include estimating page ranking, selecting top domains and documents transfer. These phases are discussed in more detail as follows:

Page ranking: page rank indicates the importance of the page by the number of times a user visits it through browsing the web giving either an implicit or explicit feedback. Implicit feedback is derived from different signals depending on the user behaviors such as the time spent on viewing documents during the search process (or dwell time), page scrolling or browsing actions [5]. On the other hand, the explicit feedback is estimated directly from the user during click-through documents as relevant documents. The feedback gives an indication of the importance related to the visited documents. The documents in the web, however, contains links to other topically related documents which forms the building block link structure for estimating the page rank of the web

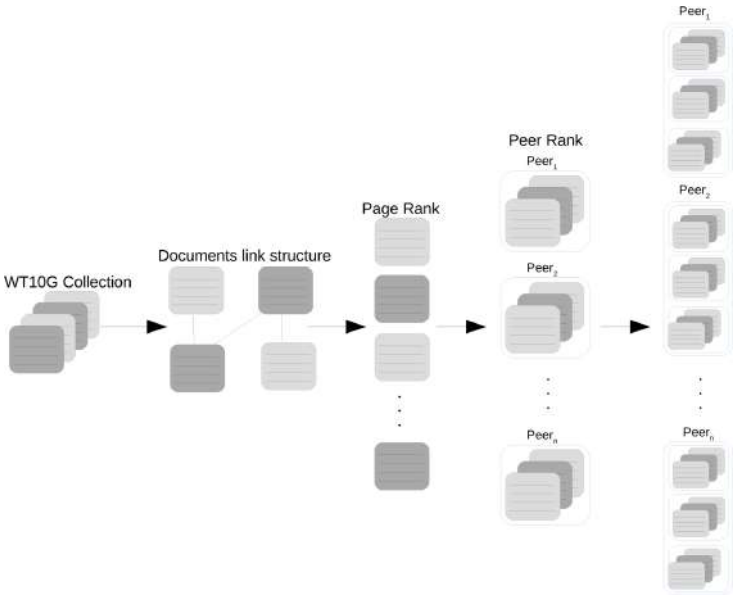

Figure 1: The process of building trust-based Testbed

pages. This rank value has the ability to distinguish the popular (those with incoming links or those that have links from popular pages) and unpopular web pages by using the in-link and out-link files [11,19]. In order to estimate the page rank of web documents, a graph of documents has to be built and traversed for derived the page rank. The graph comprises a set of nodes (or vertices) represent the documents in the corpus (or collection of documents) and a set of links (or edges) refer to the hyperlinks (in-links or out-links) between pages. Such a graph is used to estimate the document rank by walking through the graph and estimate the number of times document is visited as a probability chain the link is being visited by the others. The general formula to calculate the Page Rank for any page u is:

$$PR(u) = \frac{\lambda}{N} + (1 - \lambda) \times \sum_{v \in B_u} \frac{PR(v)}{L_v} \tag{1}$$

where N is the number of pages being considered, λ is a constant with typical value 0.15, B_u is the set of pages that point to u, and L_v is the number of outgoing links from page v (not counting duplicate links).

Peer ranking: The process of estimating the ranks of peers relies on the rank of their documents. In particular, the page ranks are estimated for each WT10g document that has in-link and out-link structures. The peer rank based on the rank of its documents is evaluated, then the peers are sorted based on their rank values in descending order. The DLPR is then built by selecting the top 8 peers from the sorted list, this number was selected to guarantee a small number of libraries and on the same time ensure the equality in the number of

documents in each peer. These top peers are used as attractors for other peers where the documents of the other peers are distributed over these attractors. This process is continued until there are no peers left in the list. As a result, this method generates 1460 number of peers each of which has 8 domains (or peers) with an average of 1160 documents.

Document transfer: The transfer of documents in the testbed ensures the uniform distribution of documents that represent the characteristics of DL applications. Nevertheless, the domain of trustworthiness needs a rank system for assigning trust values to each peer as discussed. The question is how can exploit the rank values of peers to be used as trust metric value for re-ranking the peers at query time. In summary, building a realistic testbed for DL applications reflecting the trustworthiness manner leads to the necessity for a metric that measures the trust factor of each peer collection. As a result, this paper proposes a peer-rank based metric for building a trust-based testbed.

4. Experimental setting and results

4.1 Experimental settings

Test collection. The TRECs Web 10G Collection (WT10g) is used that consists of 11,680 web domains, 1,692,096 documents and relevance assessment comprises of 100 topic query (Ad-hoc TREC 2000 task). The motivation behind using such collection is because the testbeds that are built by [9] have a set of characteristics and they are large enough to reflect the real situation in P2P architectures.

Evaluation metrics. Two metrics have used that estimate the trustworthiness values of documents and peers used in the proposed testbed that is being compared with the other two testbeds proposed by [7] and [10] as baseline testbeds. These two metrics include document ranking (or page rank) and peer ranking (extracted from the average page rank of the peers documents). Moreover, in order to evaluate the effectiveness and efficiency of the proposed trust-based testbed, a set of metrics were examined such as peer rank, the distribution of documents, fraction of peers to reach 100% Recall values and average Precision values. The peer rank value is obtained from its documents that represents the importance of a peer regarding the number of visits using the peer link structure. The distribution of the documents across peers reveals the equal number of documents over each peer to reflect the specific features characterize the DL applications. The fraction of peers required to achieve 100% Recall (for each topic) clarifies for each query how many peers in average is required to reach Recall value of 100%. The average precision represents the number of relevant documents in the results list to the whole number of documents. This metric is used in three testbeds to analyze the effect of trustworthiness on the distribution of the relevant documents over peers within the testbeds.

4.2 Experimental results and testbed anaylsis

The trustworthiness of Testbeds. The proposed trust-based testbed ranks the peers using the page rank of its documents which is estimated overall the collection documents. Here, the evaluation metrics are examined and the results are compared to the baseline testbeds. Figure 2 demonstrates the peer rank values that are estimated in the proposed testbed and the other two baseline testbeds. The x-axis represents the fraction of peers that are taken into account and the y-axis represents the average peers rank value for the fraction interval. The figure sorts out the testbeds based on the obtained average values of peers ranks across the peers fractions. The findings show that the DLPR substantially follows a power law distribution including peers rank values. This also suggests that there are in the testbed some peers are more trustworthy than the other peers and the distribution of this pattern follows the realistic DL scenario. However, such property cannot be observed for the other two testbeds, i.e. DL WOR and DL LC testbeds. As a result, these two testbeds are not suitable for DL applications depending on the trustworthiness concept to retrieve relevant and/or high quality (or trustworthy) documents from the peer-based IR systems.

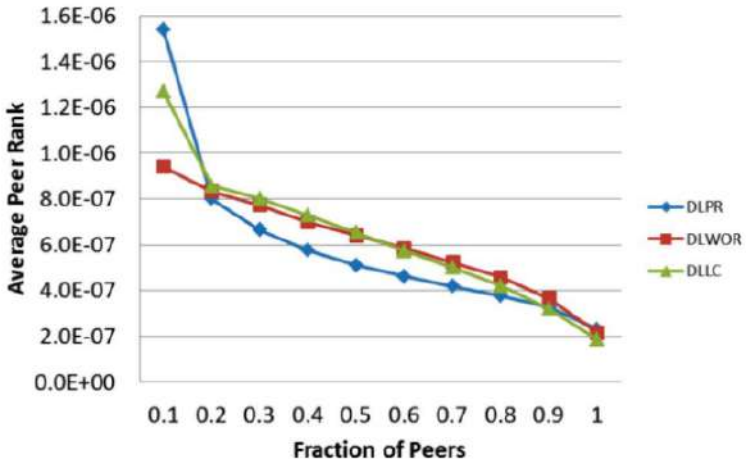

Figure 2: Average peer rank of fraction peers

Documents distribution of Testbeds. One of the P2P DLs scenario characteristics is the even distribution of documents over each digital library. Thereby, it is important to create the testbed which almost has evenly documents distribution. Figure 3 determines the average distribution of documents over fraction number of peers, where the x-axis demonstrates the fraction peers taken into account and the y-axis figures out the average number of documents in each fraction. The proposed testbed (DLPR) is the best one of the two testbeds in an equally large number of documents distribution which is more real as an initial sprinkling of documents over peer collection.

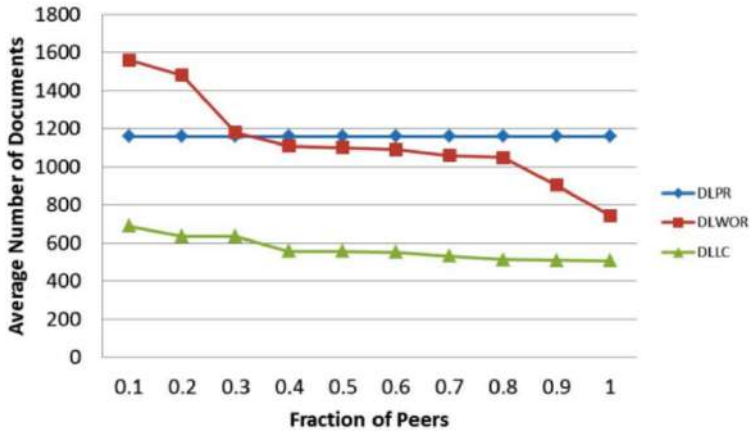

Figure 3: Average number of documents for each peer fraction

The effectiveness of Testbeds. The location of relevant documents has important attention in the P2P system architecture due to increasing the efficiency of these models. So two concepts are briefly being discussed that related to the distribution of relevant documents (topics) over peer-collection to reach 100% Recall and average precision of relevant documents over peer-collections.

Recall: In order to compute the number of peers to reach 100% Recall, the fraction of peer population is computed as shown in Figure 4. As clarified from the figure, a conclusion is that for each topic the DLPR testbeds almost has a high topical distribution over collections (or peer) than the other two testbeds. But the DLPR testbed distributes topics over peers with redundancy in an efficient manner compared to the other two baseline testbeds. As such, the probability for reaching 100% Recall is high in the DLPR over peer collections. In addition to the topics distributions, the DLPR testbed provides other important factors such as trust factors that focus on the trustworthiness of peers in providing trustworthy documents.

Precision: The other metric that is more important is the Precision within the peer-collection. The Precision represents the proportion at which topics (or relevant documents) are presented within the peer collections. The precision is calculated by using one of the [9] precision metric that considered all the peer collections have at least one relevant document and measured their average precision as follows:

$$P_{avg} = \frac{1}{n} \sum_{i=1}^{n} P_i \tag{2}$$

where n is the number of peer-collections that have at least one relevant document and P_i is the precision as measured by the number of relevant documents over the total number of documents shared in the ith collection.

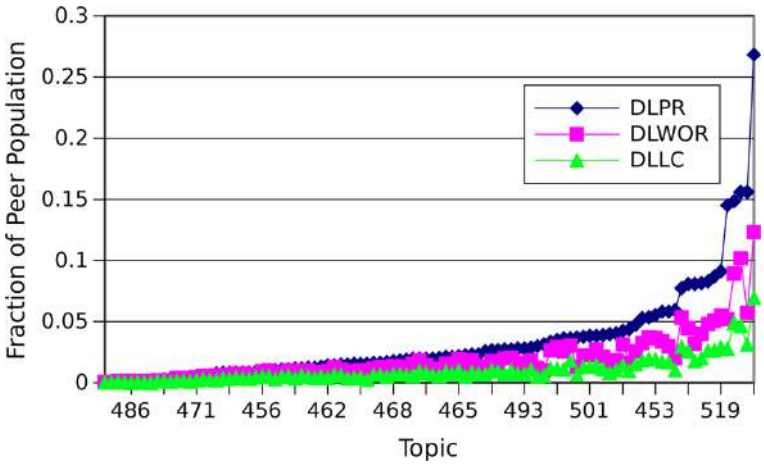

Figure 4: Fraction of peer population to reach 100% recall

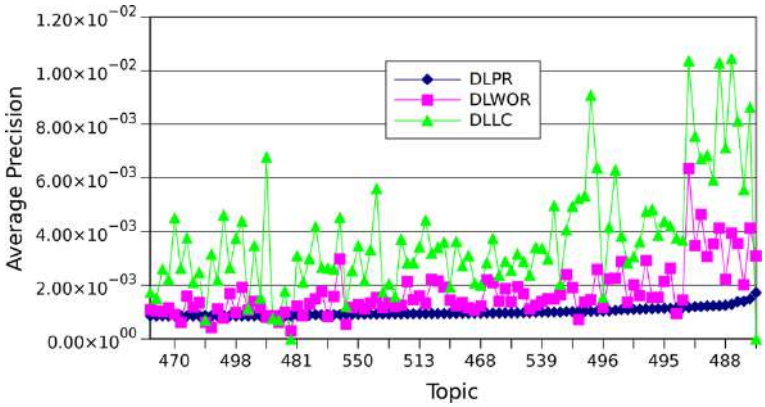

Figure 5: Average precision of topics distribution

In Figure 5, the DLLC (Digital Libraries Lu and Callan) and DLWOR (Digital Library WithOut Replication) have high Precision values than the proposed testbed. The Precision and Recall have an effect to the testbed from an effectiveness perspective, but their results do not reflect the quality (or trustworthy) of documents as the number of trustworthy documents for some peers in the testbed significantly are larger than the other two testbeds. Consequently, the DLPR testbed specifically focuses on the importance of trustworthiness that is related to some peers comparing to the other peers. The reasons behind such results back to that high number of non-relevant documents have the highest trust values, due to the trust value computed through the number of peers access to these documents or more recommend by other peers as to be the trustworthy documents [18].

5. Conclusions and future works

Building a testbed reflecting the trustworthiness factor is one of the demanding tasks. In addition, the trustworthiness concept is subjective in its nature. Therefore, estimating such factor value occurs during the lifetime of the system as deployed in dynamic systems. As a result, this paper proposed and built a testbed that comprises the trustworthiness factor required by many researchers to tackle the problem of retrieving more relevant and trustworthy (or reputable) documents. The DLPR (or proposed) testbed is constructed to reflect digital libraries by taken into considerations many factors related to them such as a large number of peers, uniformly distribution of relevant documents and the most important factor that is related to trustworthiness that depends on the page rank of the documents inside each peer. In the construction phase, a well-known dataset is selected reflecting the Web documents and the power low distribution of documents over the Internet. This dataset is the WT10g Collection that is a real collection and since it used by many other systems as a standard collection to evaluate their models.

The proposed testbed suffers from precision and recall which is expected results, as the nature of systems that depends on the trustworthiness changes content through the lifetime of systems and some peers might join and leaf the system to increase their reputation and existence so the distribution of topics on the type of these systems was not taken into consideration.

Evaluating the trustworthiness in the P2P overlay networks is one of the most demanding and neglected tasks because the trustworthiness is subjective in nature and it supposed to be computed through the lifetime of the system. This paper presented a proposed trust-based testbed that depends on the peer rank as a trust factor used as a criterion for building the testbed. The distribution of peers ranks in peer collections of DLPR, DLWOR and DLLC testbeds is also discussed along with demonstrating the other factors such as the distribution of relevant documents (presented by topics) over peers collection in order to reach 100% recall and the distribution of relevant documents (presented by topics) over peers. The proposed testbed is concentrated on the importance of peer on peer-collection and reflects it as peer trustworthiness, so another factor may be relatively different. In the future work, the proposed testbed is being expanded to support the model with the actual behavior of a P2P system to check how much accurate the testbed in trust-based P2P systems.

References

[1] R.S. Alkhawaldeh, P. Deepak, J.M. Jose, F. Yuan, *Ltro: learning to route queries in clustered p2p ir*, In European Conference on Information Retrieval, Springer, 2017, 513519.

[2] R.S. Alkhawaldeh, J.M. Jose, *Evaluating document retrieval methods for resource selection in clustered p2p ir*, In Proceedings of the 25th ACM

International on Conference on Information and Knowledge Management, CIKM'16, New York, NY, USA, 2016, 20732076,

[3] R.S. Alkhawaldeh, *Query routing in cooperative semi-structured peer-to-peer information retrieval networks*, PhD thesis, University of Glasgow, 2016.

[4] B.R. Cami, H. Hassanpour, *A reputation-based trust model with fuzzy approach and dp,q-distance technique for peer-to-peer networks*, IJCA, 37 (2012), 4144.

[5] T. Joachims, L. Granka, B. Pan, H. Hembrooke, G. Gay, *Accurately interpreting clickthrough data as implicit feedback*, In ACM SIGIR Forum, 51 (2017), 411.

[6] S.D. Kamvar, M.T. Schlosser, H. Garcia-Molina, *The eigentrust algorithm for reputation management in p2p networks*, In Proceedings of the 12th International Conference on World Wide Web, WWW'03, New York, NY, USA, 2003, 640651.

[7] I.A. Klampanos, J.M. Jose, *An evaluation of a cluster-based architecture for peer-to-peer information retrieval*, In DEXA, 2007, 380391.

[8] I.A. Klampanos, J.M. Jose, *Searching in peer-to-peer networks*, Computer Science Review, 6 (2012), 161183.

[9] I.A. Klampanos, V. Poznanski, J.M. Jose, P. Dickman, E.H. Road, *A suite of testbeds for the realistic evaluation of peer-to-peer information retrieval systems*, In ECIR, 2005, 3851.

[10] J. Lu, J. Callan, *Content-based retrieval in hybrid peer-to-peer networks*, In CIKM, 2003, 199206.

[11] F.A. Massucci, D. Docampo, *Measuring the academic reputation through citation networks via pagerank*, Journal of Informetrics, 13 (2019), 185201.

[12] S. Sawamura, A. Aikebaier, T. Enokido, V. Barolli, M. Takizawa, *Trustworthiness of acquaintances in peer-to-peer(p2p) overlay networks*, In CISIS, 2010, 267274.

[13] M. Swan, *Blockchain: blueprint for a new economy*, O'Reilly Media, Inc., 2015

[14] A.S. Tigelaar, D. Hiemstra, D. Trieschnigg, *Peer-to-peer information retrieval: an overview*, ACM Trans. Inf. Syst., 30 (2012), 19.

[15] Y.B. Udupi, P. Yolum, M.P. Singh, *Trustworthy service caching: cooperative search in p2p information systems*, 2003.

[16] G. Wan, Z. Liu, *Content-based information retrieval and digital libraries*, ITAL, 27 (2008), 41.

[17] L. Xiong, L. Liu, *Peertrust: a trust mechanism for an open peer-to-peer information system*, Technical report, 2002.

[18] Y. Zhang, *Trust-aware information retrieval in peer-to-peer environments*, PhD thesis, The University of Manchester, Computer Science, 2011.

[19] Z. Zhang, J. Gao, F. Ciravegna, *Semre-rank: improving automatic term extraction by incorporating semantic relatedness with personalised pagerank*, ACM Transactions on Knowledge Discovery from Data (TKDD), 12 (2018), 57.

[20] H. Zhao, X. Li, *Vectortrust: trust vector aggregation scheme for trust management in peer-to-peer networks*, In 2009 Proceedings of 18th International Conference on Computer Communications and Networks, 2009, 16.

Accepted: 16.01.2019