A Survey and Evaluation of State-of-the-Art Intelligent Question Routing Systems

Bojan Furlan, Bosko Nikolic, and Veljko Milutinovic, Fellow of the IEEE

School of Electrical Engineering, University of Belgrade, Serbia

{bojan.furlan, bosko.nikolic, veljko.milutinovic}@etf.bg.ac.rs

Abstract – In spite of great developments in artificial intelligence, human brain is still more powerful, concerning the comprehension and manipulation with partially known facts. One domain where this is prominent is related to a problem of question answering. When it comes to giving the answers, especially those that do not explicitly exist in the text corpus, the advantages of a human expert are abilities like explaining, combining complex answers, and abstract reasoning. This paper represents a survey of the existing research in the domain of intelligent question routing. The survey starts from an original presentation paradigm that generalizes the essence of approaches found in the open literature. The presentation paradigm includes three basic processing stages related to the three major problems of system implementation: question analysis, question forwarding, and users' knowledge profiling. Various research efforts use different approaches for implementation of each one of the basic processing stages. Each particular approach is described here using the same template. All these approaches are enlisted, discussed, and presented within a table, for easier comparison. The outcome of this analysis is a proposal for a new approach that tackles identified problems.

I. Introduction

The basic idea of Intelligent Question Routing Systems (IQRS) can be informally stated as: do not ask machines to understand humans (as they are not capable of it, at this level of technological development), but ask them something easier, to find a person that is able to give the correct answer. Therefore, if there is a need for an advice or a direction, instead of seeking through a huge amount of information that Internet is surely offering, another option is to ask a person who is competent to provide a short and human understandable answer. As a result, intellectual task is still left to a human, but the burden of finding the right person for particular question is delegated to a machine.

Given that IQRS aims to serve as a knowledge exchange medium in an arbitrary field of expertise, the goal of this survey paper is to shed light on a plethora of novel approaches. These approaches are of importance for a number of applications, where intensive communication between users is required (e.g., large enterprises, e-government agencies, technical support, health care system, army, etc.). Other applications can involve a support in educational and collaboration processes, where IQRS facilitates an efficient and effective knowledge exchange between scientists, researchers, university staff, and students. The benefit coming from deployment of such systems includes: (a) reducing unnecessary "pinging" of experts, which are a valuable resource and (b) increasing the system owners' (enterprise, government, university) quality of service, since users are more satisfied with answers, because their questions are answered by the right persons.

Figure 1 illustrates a typical IQRS scenario that consists of the following 6 phases: (1) type a question, (2) understand "information need," (3) find a right person, (4) forward the question, (5) understand the question, and (6) give an answer. In the first phase, a user with a question – asker, types the question and sends it to the IQRS service. Then, the service has to understand "information need" from the question, find a right person (potential

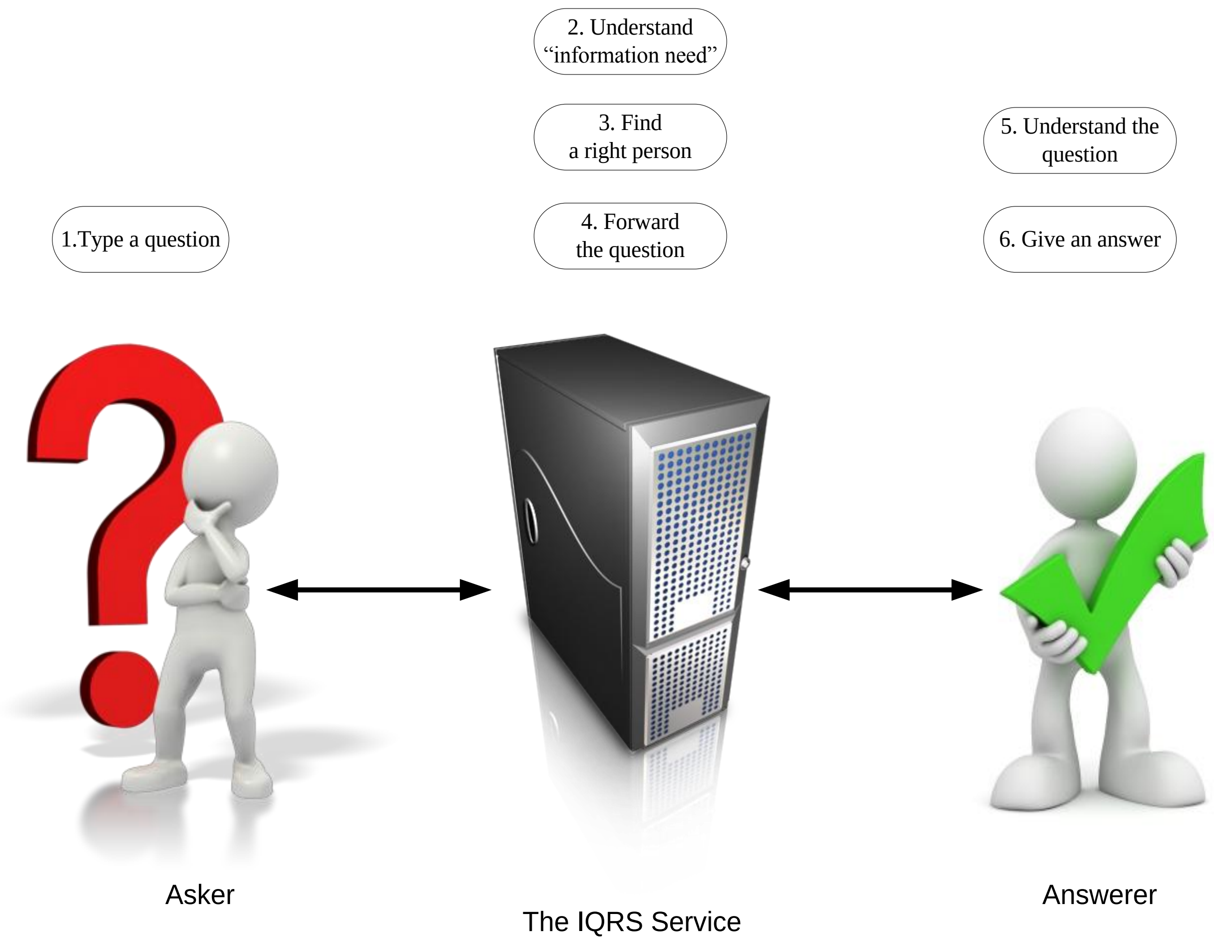

Figure 1. Illustration of an IQRS task: A typical scenario

answerer), and then forward it to that person. Finally, the rest of the answering task is left on the human. When potential answerer receives the question, after reading and understanding it, he or she has an option of giving an answer. Phases (2), (3), and (4) are delegated to the IQRS service, thus they will be the main focus of this survey.

The rest of the paper is structured as following: it starts from a discussion on how the IQRS domain of research is related to other similar fields and why it is important. Then, it gives an original presentation paradigm that generalizes the essence of approaches found in the open literature. After that, each particular approach is described using the same template. All these approaches are enlisted, discussed, and presented within a table, for easier comparison. Finally, the outcome of this analysis is given as a proposal for a new approach based on a generalized treatment of the user knowledge profiling.

II. Background

Question Answering (Q/A) is the task of automatic answering a question posed in a natural language. The main purpose of Q/A systems was, and still is, to move the retrieval focus from Document Retrieval to Information Retrieval, by extracting relevant and concise answers to a wide range of open domain questions. For finding the answer, Q/A systems use diverse data sources from pre-structured databases to large collections of documents written in a natural language (text corpus). This text corpus can consist of formal documents like compiled newswire reports[1], or from noisier ones (and not strictly formatted) such as blogs from the World Wide Web[2]. However, sometimes it is necessary to adopt fundamental knowledge from a variety of domains. On the contrary, quite many experts with needed knowledge exist within some institution, company, or university. For a young researcher, a student or someone curious about a new topic, it would be very helpful to contact directly a person competent in a particular domain and ask him/her for an advice or instruction. The efficiency of finding the right person can be gained by using a software system for intelligent question routing.

Also, the IQRS task is related to, but distinct from, link analysis and expert finding. Link-based algorithms like PageRank and HITS were successfully applied in social media to find experts[3]. However, previous approaches could not tackle the question routing problem since they are only considered from the user's perspective; but if a particular question is presented, they could not utilize the specific characteristic of that question to determine whether a user would answer it. Finally, an extensive research has been conducted in the field of Semantic Query Routing (SQR) in peer-to-peer networks. One example of how to locate peers that are relevant with respect to a given query is by building a semantic overlay network[4]. Queries are routed through a super-peer where every peer needs to explicitly advertise its content. Another example is implicit content identification based on social metaphors[5]. However, in IQRS we refer to questions as a free-form text, as opposed to structured or semi-structured queries. Therefore, with advancement of text processing tools and a recent boom of social networks, the synergy of Q/A, expert finding, and SQR has become possible, so question routing between users (IQRS) is an open research issue. Accordingly, the rest of the paper represents a survey of related work focused only on IQRS and related problems of importance for the paradigm introduced in this paper.

III. GENERALIZATION OF ANALYZED APPROACHES

Authors assume that question routing is a complex process influenced by both static and dynamic parameters, so the results of the presented research are more widely applicable. The viewpoint of this survey is best represented by the notions of Figure 2. Since the question-answering task consists of "asking the question" and "giving the answer," the presented structure is divided in two parts that simultaneously process: (a) new questions

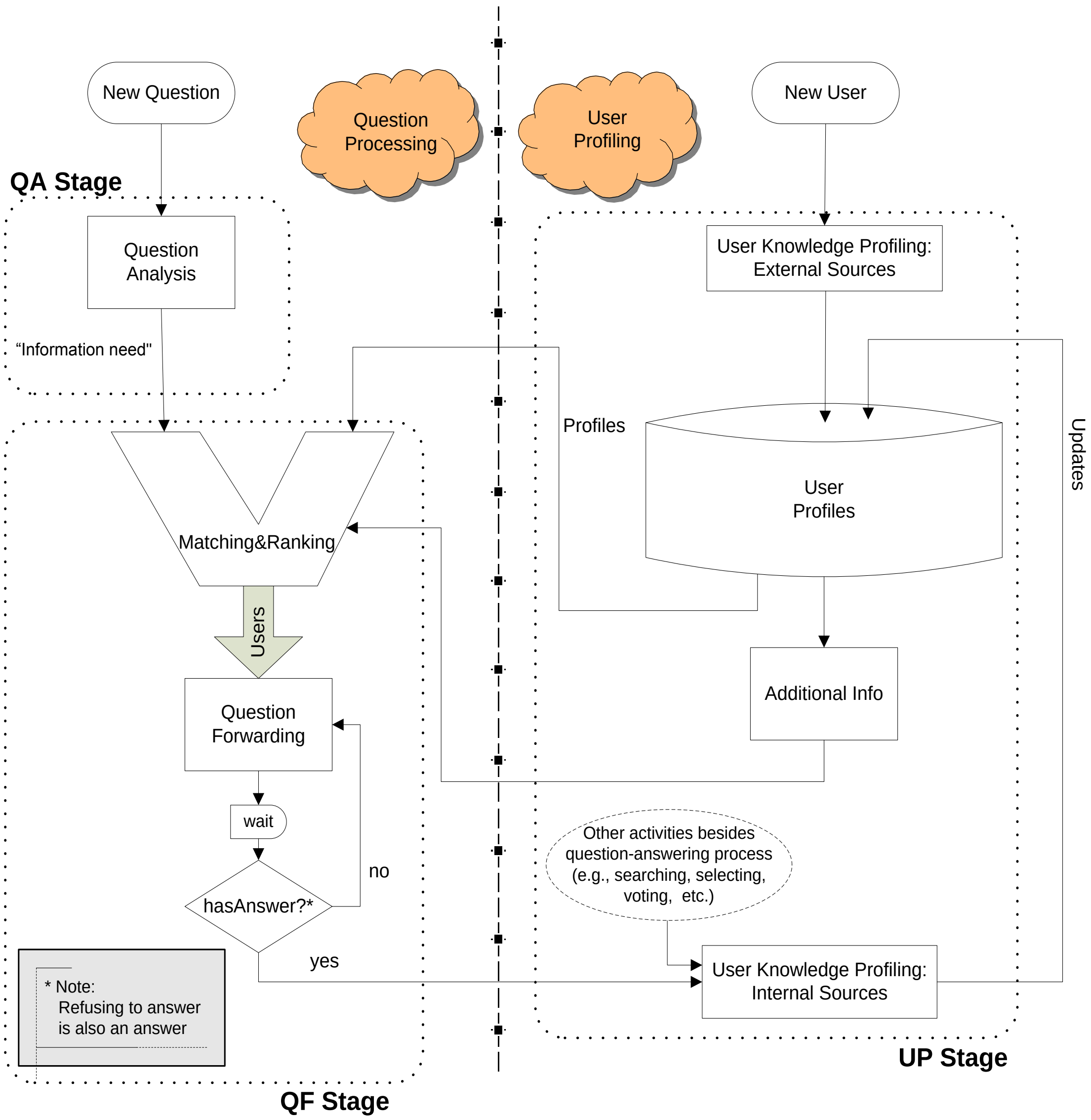

Figure 2. An Original Presentation Paradigm: Generalization of Approaches from the Open Literature

(Question Processing) and (b) new or existing users (User Profiling). Both parts consist of stages represented in Figure 2. Each stage contains one or more modules, which are implemented using an algorithm from a relatively large pool of algorithms. Question Analysis (QA) Stage contains only *Question Analysis* module, Question Forwarding (QF) Stage contains *Matching&Ranking* and *Question Forwarding* modules, and User Profiling (UP) Stage contains the following modules: *User Knowledge Profiling: External Sources*, *Additional Info*, and *User Knowledge Profiling: Internal Sources*, as well as the *User Profiles* repository.

Each one of these stages is related to one of the three main issues that are defined by the three questions elaborated in the text to follow:

Question#1: *How to identify information need from a question?*

In IQRS, the requirement for a question analysis is only to be able to understand the question sufficiently for routing it to a competent answerer. This is a considerably simpler task than the challenge facing an ideal Q/A system, which must attempt to determine exactly what piece of information the user is seeking (e.g., to translate information need into search keywords), to evaluate whether a founded content includes that piece of information, and to extract it to a human-understandable format. By contrast, in IQRS, it is the human answerer who has the responsibility for determining the relevance of an answer to a question, and that is a function which human intelligence is well-suited to perform.

The task of question analysis and information need extraction is presented by the *Question Analysis* module. Usually this module extracts some relevant terms/concepts from a question or classifies a question into one or more afore-determined topics. Therefore, the output that is in a form of this identified terms or topics is then forwarded to the *Matching&Ranking* module.

Criteria of interest for **Question#1**:

1) User interaction type:
    a) With question annotation (e.g., tags or predefined categories),
    b) Without question annotation.
2) Algorithm extraction type:
    a) Natural Language Processing techniques - NLP (e.g., stemming, Part Of Speech - POS processing and filtering, synonym lookup),
    b) Data Mining/Machine Learning techniques - DM (e.g., trained topic classifiers) or ML (topic modeling).

Possible improvement avenues for **Question#1:** Question visualization.

Question#2: *How to find competent users for a particular question?*

Given the information about the question derived from the question-analysis module, the task of finding competent users is performed by matching the recognized information need from the question against the available knowledge profiles, and ranking them in an ordered list of users (or "candidate answerers") who should be contacted to answer the question. This matching can be realized as exact matching or by computing a semantic similarity. Also, the model of this process can be organized as centralized (matching and ranking is realized within one central node), or can be distributed (all nodes are employed within this process).

As shown in Figure 2, the inputs of the *Matching&Ranking* module are: (a) an information need recognized form the question, (b) available knowledge profiles from the user profiles repository, and (c) additional info like availability, response rate, or popularity rank. The output is an ordered list of users that is submitted to the *Question Forwarding* module.

Criteria of interest for **Question#2**:

1) Model organization
   a) Centralized
   b) Distributed
2) Similarity calculation
   a) With exact matching
   b) With semantic matching

Possible improvement avenues for **Question#2:** Incorporation of semantic and string similarity.

Question#3: *How to accurately profile user's knowledge from various information sources?*

Knowledge can be classified broadly as either explicit or tacit[6,7]. Explicit knowledge consists of facts, rules, relationships, and policies that can be faithfully codified in paper or electronic form. Since it is explicitly expressed, it can be shared without a need for discussion. By contrast, tacit knowledge (or intuition) requires interaction. This kind of knowledge underlines personal skill, and is largely influenced by beliefs, perspectives, and values. Its transfer requires face-to-face contact or even apprenticeship. Since individual knowledge is learned (internalized) into the human brain, the psychological approach by observing the subject's characteristics from the performed behavior has to be applied. In this case, the observed behavior is represented by the content that a user generates. To some extent this content maps to explicit and tacit classification of knowledge, where explicit knowledge is mostly expressed within the published documents, such as scientific papers, books, articles or blogs, while email communication and content generated during the question-answering process can identify the tacit knowledge. As a result, both sorts of information are valuable for profiling of user knowledge.

The IQRS keeps user profiles in a repository that is constantly being updated. Besides expertise, which is the prime information kept about a user, the user profile can also contain aforementioned additional information (*Additional Info*). This information is not directly related to knowledge, but can improve system's quality of service (e.g., responsiveness if questions are preferably routed to users with high response rate). On the other hand, part of the user profile related to expertise can be built from various information sources. These sources can be classified into two categories: (a) internal and (b) external.

a.) Internal sources capture information about user activities within IQRS, some directly related to Q/A process like asking or answering, or some implicit like searching, selecting, voting, etc. Further, this can be divided regarding employed profiling technique into two subcategories: text (e.g., for information extracted from question/answer threads) and other (e.g., for information extracted from question-reply links between users or response scores).

b.) External sources are usually exploited for initial user knowledge profiling in order to avoid the new user (cold-start) problem, as well as for following profile updates. They capture information from other systems that are not an integral part of IQRS, such as social networks, blogs, email repositories, etc. Also, these sources can be divided into the same two subcategories: text (e.g., posts on social networks, blogs, etc.) and other (e.g., demographic and behavior information, social connections, etc.).

As represented in Figure 2, for a new user, the initial profile is created in the *User Knowledge Profiling: External Sources* module. Afterwards, updates about user are gathered from the question-answering process, from the *User Knowledge Profiling: Internal Sources* module (e.g., correct or incorrect answers rates) or from some external updates, from the *User Knowledge Profiling: External Sources* module (e.g., manually changing his/her profile).

Criteria of interest for **Question#3:** user profiling methodology by expertise identification:

1) Text:

   a) Natural Language Processing techniques - NLP (e.g., stemming, ad-hoc named entity extractor),

b) Data Mining/Machine Learning techniques - DM (e.g., classification, clustering) or ML (topic modeling),

c) Recommender System (RS) model.

2) Other:

a) ad-hoc (AH) model,

b) Recommender System (RS) model,

c) Data Mining – DM (e.g., PageRank or HITS).

Possible improvement avenues for **Question#3**: Profile integration.

IV. Presentation of Analyzed Approaches

Approaches presented in the text to follow address the three main issues that define IQRS in a characteristic manner. For easier comparison, all approaches are also presented in Table1. Each column in the table corresponds to a particular element of the intelligent question routing process from Figure 2. Each table entry includes the

TABLE I. Comparison of selected Approaches

	1. Question Analysis		**2. Matching & Ranking**		**3. User Knowledge Profiling**				**4. Additional Info.**
	Annotation	***Analysis***	***Model Organization***	***Semantic Matching***	***Internal***		***External***		
					Text	***Other***	***Text***	***Other***	
A. iLink	Tagging	NLP	Centralized (or Distributed)	No	DM	Response Score	DM	Manual	Referral Rank
B. PLSA in CQA	No	ML	Centralized	No	RS & ML	No	No		No
C. Routing within Forums	No	NLP	Centralized	Yes	ML	DM	No		Authority
D. Question Routing Framework	No	No	Centralized	No	RS & ML	No	No		Availability
E. G-Finder	No	Heuristics	Centralized	Yes	DM	DM	No		Concept Network
F. Aardvark	Tagging	DM	Centralized	Yes	DM & NLP	RS	DM & NLP	RS & Manual	Connectedness Availability
G. Confucius	Categories	DM	Centralized	Yes	DM	DM	No		Timeliness, Coverage, and Spam
H. Yahoo! Answers Recommender System	Categories	NLP	Centralized	No	RS	RS	No		Group of User Attributes
I. STM in CQA	Tagging	NLP	Centralized	Yes	ML	No	No		No
J. Classification-based Routing in CQA	Categories	NLP	Centralized	Yes	DM	DM	No		Global Features
K. SQM	No	No	Distributed	No	No	Expertise Score	No		Response Rate

name and a short description (within the criteria of interest). Also, all analyzed approaches are described in a similar way, including the information according to the following template: essence, structure, relevant details, applications, and pros & cons.

A. iLink

Davitz J. et al[8] in 2007 proposed a model for social search and message routing named iLink. They focused on the problem how to model social networks and how those networks accomplish tasks through peer-to-peer production style collaboration. The iLink model has been used to develop a system for generation of Frequently Asked Questions (FAQ) in a social network - FAQtory. The system facilitates generation of a repository of question/answer threads, so when users send questions to the system they are presented with a list of related question/answer pairs, a list of experts on the topics found in the question, and as a last resort, search results from the web. The social network is represented as a graph with nodes and links. Expertise, response score, and referral rank are maintained for every node (user). NLP is used for *Question Analysis*, particularly stemming, synonym lookup, and stop word removals. Users are also allowed to tag questions in order to improve system's performance. DM technique (clustering) is used for *User Knowledge Profiling* from text sources. Other maintained parameter is response score, which is in function of response rate, response accuracy, etc. The *Matching&Ranking* module does not employ semantic similarity between terms. Referral rank about the user is maintained as *Additional Info*, which correlates to popularity referrals from other users. Clustering techniques are employed for initial user knowledge profiling that extract information from external sources, such as an underlying social network, email repository, or user's homepage. Also, users are allowed to do manual profile edits. The iLink model organization is centralized (as a supernode in the social network), but it can be also used in a decentralized manner. The structure is illustrated in Figure 3 using the template introduced in the generalization paradigm.

An interesting idea introduced with iLink, is that it allows incremental answering. At each step in a query thread, user nodes can contribute some information, even if that information does not qualify as an answer. This information can be about the query itself or it can simply be some evidence about where knowledge might exist in the network (e.g., who knows something, who knows somebody). On the other hand, the iLink does not use general semantic similarity matching, since it requires some prior knowledge about relevant topics in the domain. It also lacks of noise robustness and some more detailed incentive model.

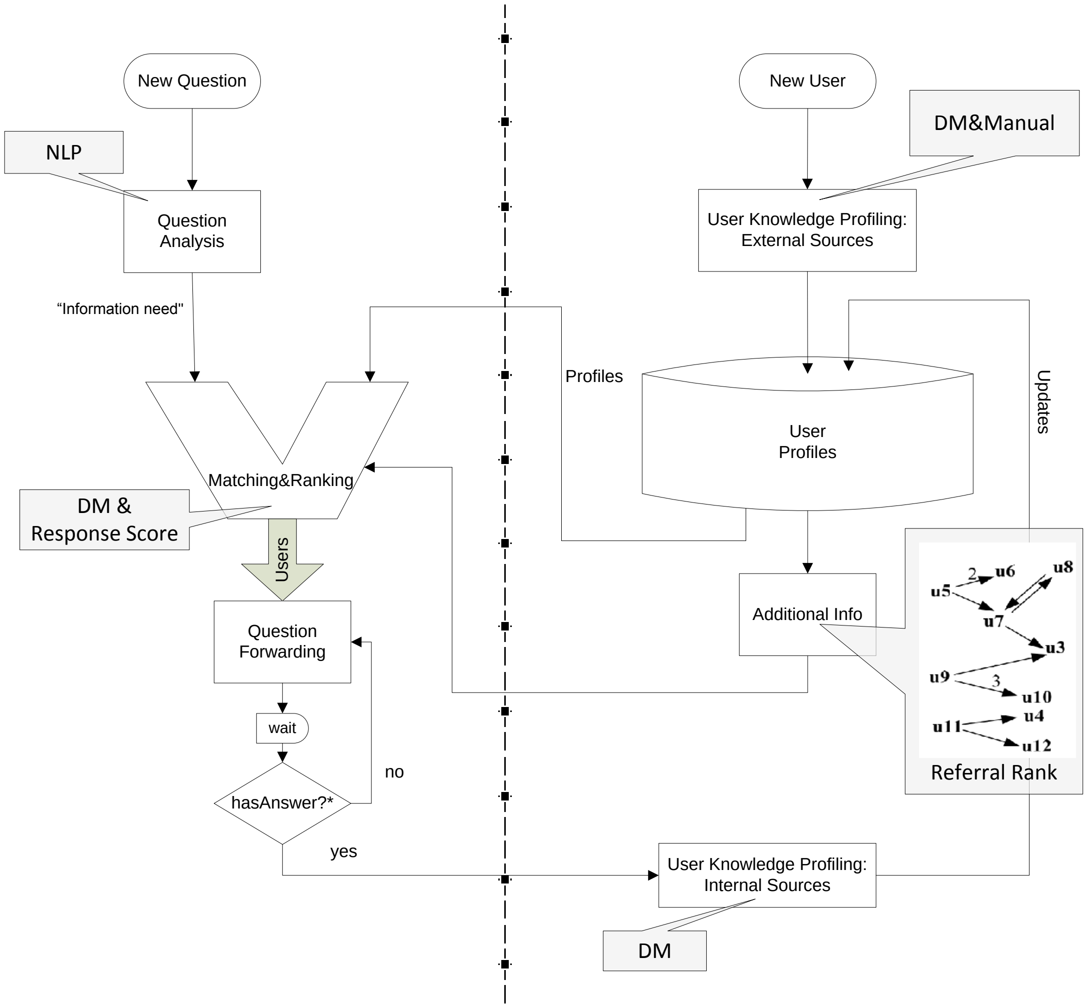

Figure 3. The iLink: Structure

B. Probabilistic Latent Semantic Analysis in Community Question Answering Portals

Qu M. et al[9] in 2009 proposed a question recommendation technique using the Probabilistic Latent Semantic Analysis (PLSA) that helps users to locate interesting questions in Community Question Answering (CQA) portals such as Yahoo! Answers. The PLSA topic modeling technique is used for *User Knowledge Profiling* from text sources. Also, for *Question Analysis* the same machine learning technique (PLSA) is used. *Matching&Ranking* is centralized and it is not using semantic similarity match between extracted terms. The structure of this model is illustrated in Figure 4.

This paper is included in the survey since it introduced an innovative approach to knowledge profiling based on the topic modeling technique. Also, it proposed a novel metric to evaluate the approach performance by matching a recommended user's rank with the best answerer's rank in Yahoo! Answers dataset. However, it lacks of many analyzed attributes. There is no possibility of question annotation and no *Additional Info* about user is maintained. Also, there are no other parameters incorporated in the knowledge profile.

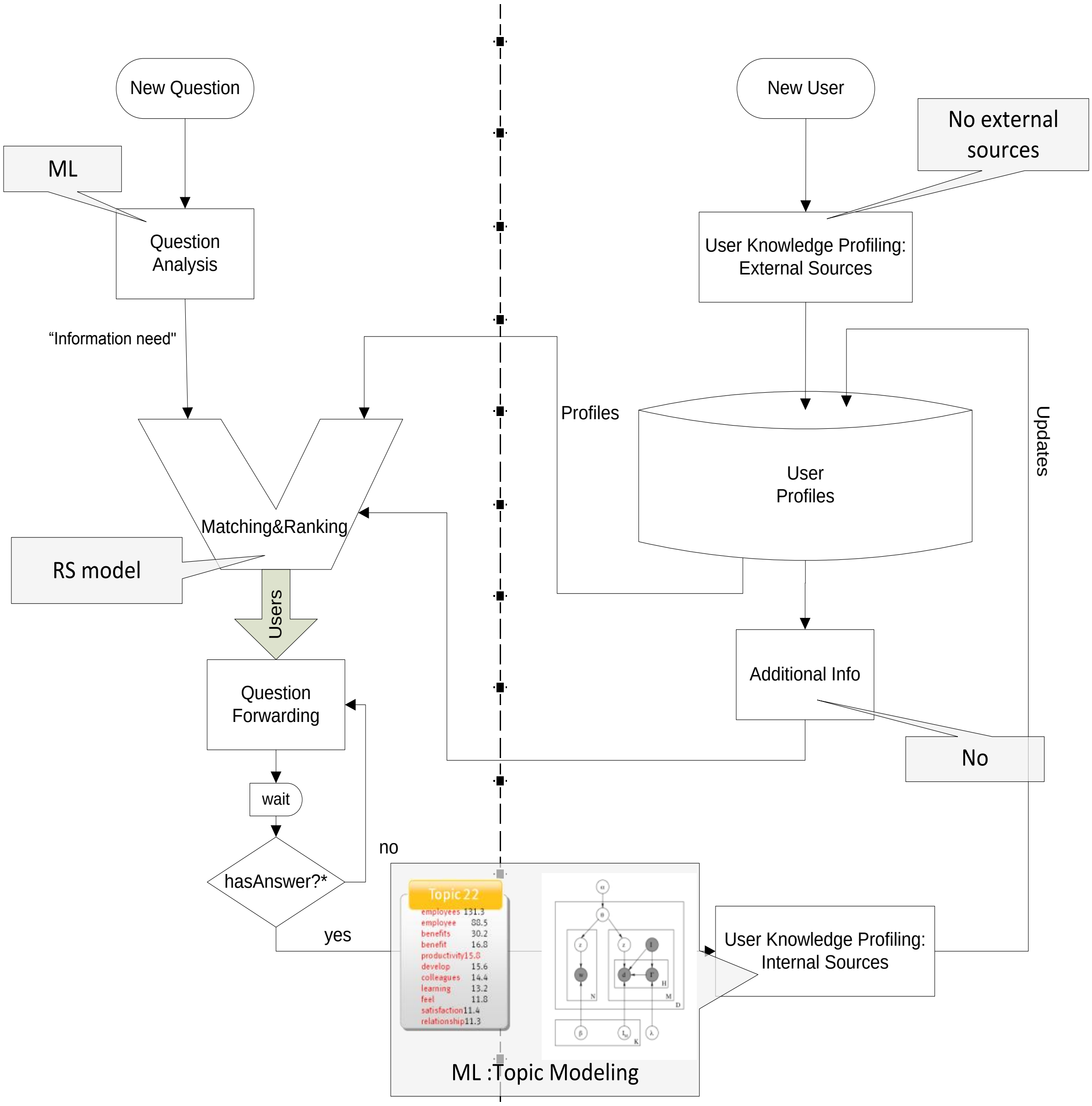

Figure 4. The PLSA in CQA: Structure

C. *Routing within Forums*

Zhou Y. et al[10] in 2009 proposed a combination of the language model and the post-reply structure of forum threads to route questions to proper users. Machine learning techniques based on term probabilities are employed to profile the forum participants (*User Knowledge Profiling*). More precisely, three different policies were used to integrate three different language models: the profile-based model, the thread-based model, and the cluster-based model. The profile-based model creates a profile for each user based on the answers authored by the user and also the corresponding questions that he/she answered. In the thread-based model each thread serves as a latent topic, so the probability of a user being an expert for a new question is computed based on each thread and the association between the thread and relevant users. Therefore, each thread-based profile contributes to the ranking score of a user on the basis of the association of the thread with users. The cluster-based model groups threads with similar content into clusters and builds a cluster-based thread for each user. Each cluster represents a coherent topic and is associated with users that indicate the relevance between the cluster and users. Given a new question, the ranking score is then computed for each user by aggregating all clusters. To avoid the zero probability for unseen words all three language models are smoothed with a background language model, so a semantic matching is implicitly employed. Questions as well as answers within threads are pre-processed, which includes tokenization, stop words filtering, and stemming. Simultaneously with the language models, a global ranking of users is employed using the authority values, which are computed by the PageRank-based algorithm on the post-reply graph. User profiles are then combined with this authority values to produce the final ranking of experts. Therefore, a final ranking score for each user is computed from a probabilistic model that integrates the results of these two steps – language model based profiling and authority based re-ranking. The structure of the system is presented in Figure 5.

Evaluation was performed on a thread data collected from Tripadvisor forums. Results showed that employed algorithms made significant improvement on precision and recall. On the other hand, a problem of updates

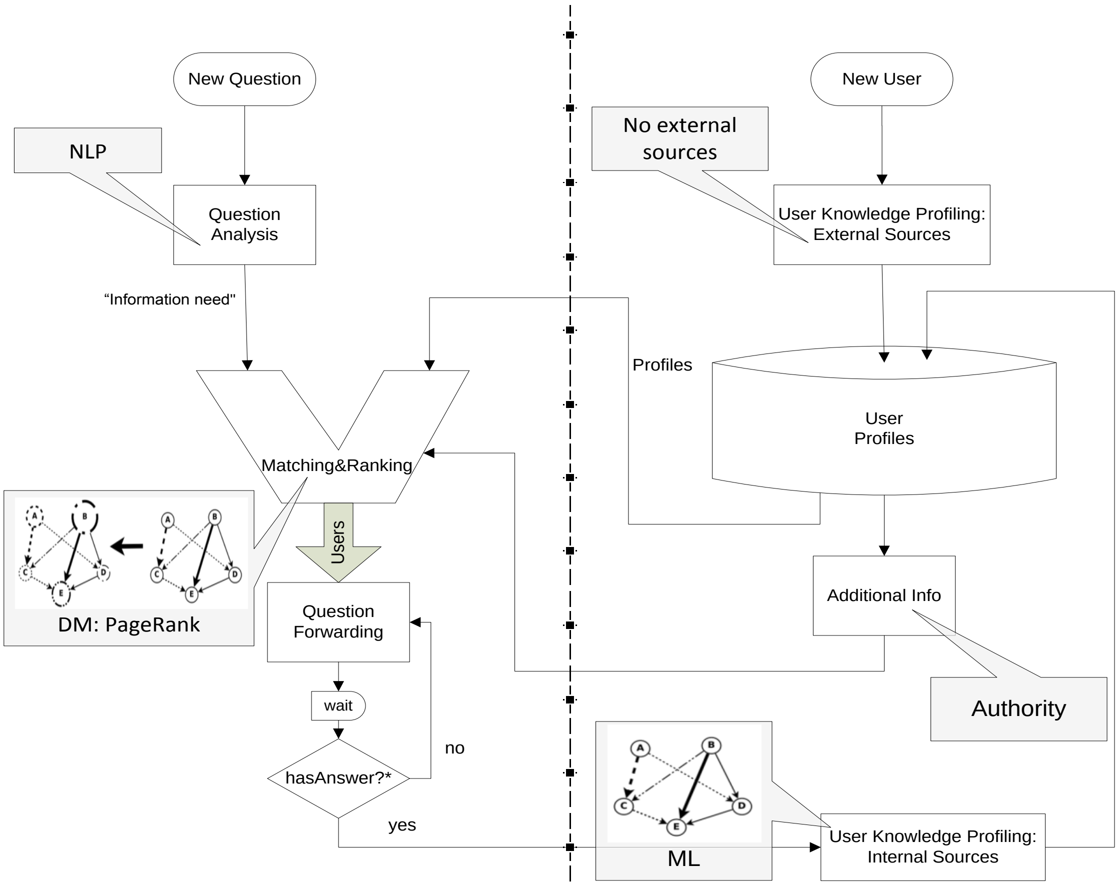

Figure 5. Routing within Forums: Structure

remains since new threads are posted everyday in forums, creating a need to update the inverted indexes with new forum threads. This does not refer to incremental update of the thread-based model, while it appears to be nontrivial for the profile-based and cluster-based models. Also, some clustering techniques can be applied to generate fine grained clusters instead of using the clusters generated by sub-forums. Finally, the new user problem exists since no external sources are used for creating initial profile.

D. Question Routing Framework

Li B. and King I.[11] in 2010 proposed a framework called Question Routing (QR) that ranks the answerers in CQA. *User Knowledge Profiling* from text is done twofold: with and without consideration of an answer quality. The first estimation models the potential answer quality based on the quality of previously answered questions by the user. The second one uses only term frequencies for calculating similarity between a particular question and all previously answered questions. Also, availability is estimated as *Additional Info*. It is assumed that a user is available to provide answers for the routed questions when he/she is logged on the system, so estimation is made by modeling this problem as a trend analysis problem in time-series data mining. *Matching&Ranking* is centralized and for each potential answerer the final QR score is calculated as a linear combination of the estimated expertise score and the availability score. The structure of the framework is illustrated in Figure 6.

The QR framework considers both users' expertise and users' availabilities for providing answers in a range of time. Conducted experiments with Yahoo! Answers dataset demonstrated an improvement by applying the proposed framework. Still, since no external sources are used for creating an initial profile the new user problem exists. Also, incorporation of other attributes in user profiles is not considered, as well as semantic matching between extracted terms. Finally, there is no question analysis or question annotation.

E. G-Finder

Li W. et al[12] in 2010 presented their design and implementation of G-Finder, an algorithm and a tool that makes intelligent routing decisions within programming forums (e.g., Java programming forums). The work was motivated by the empirical study conducted over three popular programming forums, which showed that the forum users experience a long waiting period for answers and a small number of experts are often overloaded with questions. Therefore, their goal was to leverage the source code information of the software systems that forums are dedicated to, and discover latent relationships between forums users. *User Knowledge Profiling* is carried by two algorithms that exploit the program source code and the forum data to construct the concept networks and the user networks and integrate them into a probabilistic model. Concept networks (e.g., Java class

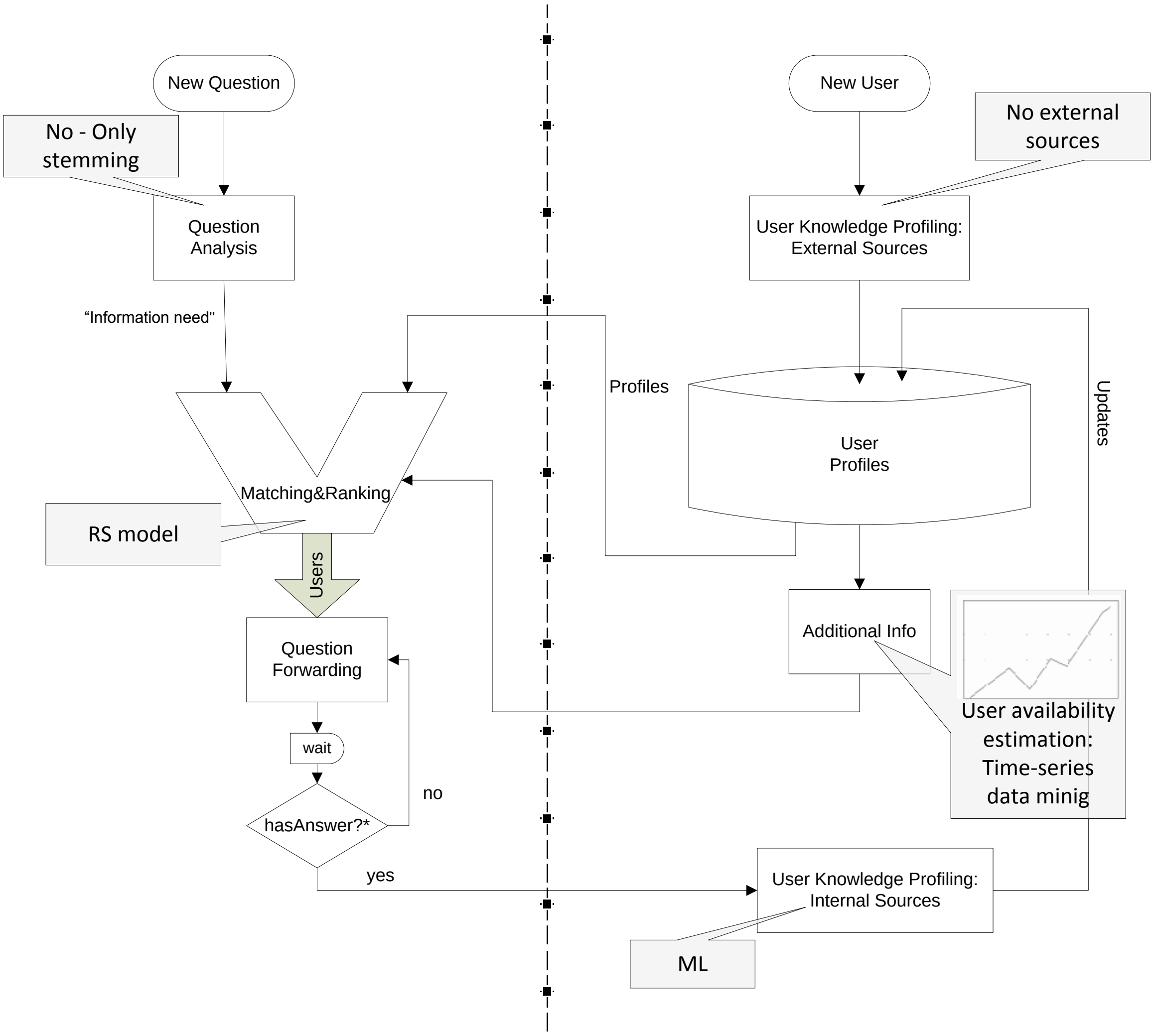

Figure 6. The Question Routing Framework: Structure

types) are dynamically built by extracting the concepts relationships (e.g., class type hierarchies or a method call graph), either from the source code or from the bytecode of the system that the forum is dedicated to. Simultaneously, user networks are constructed as directed graphs from the post-reply relationships. Programming questions are then used to integrate these two networks and rank the potential experts (*Matching&Ranking*) by calculating:

$$\mathrm{P}(\mathrm{user}|\mathrm{question}) = \sum_{concept} \mathrm{P}(\mathrm{user}|\mathrm{concept}) \times \mathrm{P}(\mathrm{concept}|\mathrm{question})$$

The P(user|concept) probability is computed through two steps:

(1) Construct the user network on each concept and apply adjusted PageRank algorithm. In this step, the relationships of concepts are not involved.

(2) Take into consideration the relationships among concepts, represented by the concept network, and for the user compute the P(user|concept) probability based on the semantic clustering of concepts.

Finally, P(concept|question) relates to *Question Analysis* that is performed by using heuristics, such as concept occurrence within a question title or a concept TF/IDF metric within a source code of a message body. The structure of G-Finder is presented in Figure 7.

The evaluation carried out by using the data from three large programming forums (the Java Forum, the Java DevShed Forum, and the GEF Forum) showed that G-Finder considerably improves the prediction precision of experts who can provide answers to programming questions. A major limitation is the heuristic-based concept mapping method that in some cases fails to extract concepts from threads or questions. Also, some latent patterns and relationships that are not captured by the model are the new user (expert) problem, the general expert problem and the random expert problem. New user problem relates to users who never reply to anything before so their expertise cannot be profiled by the model. Manifest of general expert problem is that questions answered by this type of users have no explicit or implicit semantic relationships with each other. The random expert problem relates to users that actively participate in the threads with comments and suggestions but rarely with

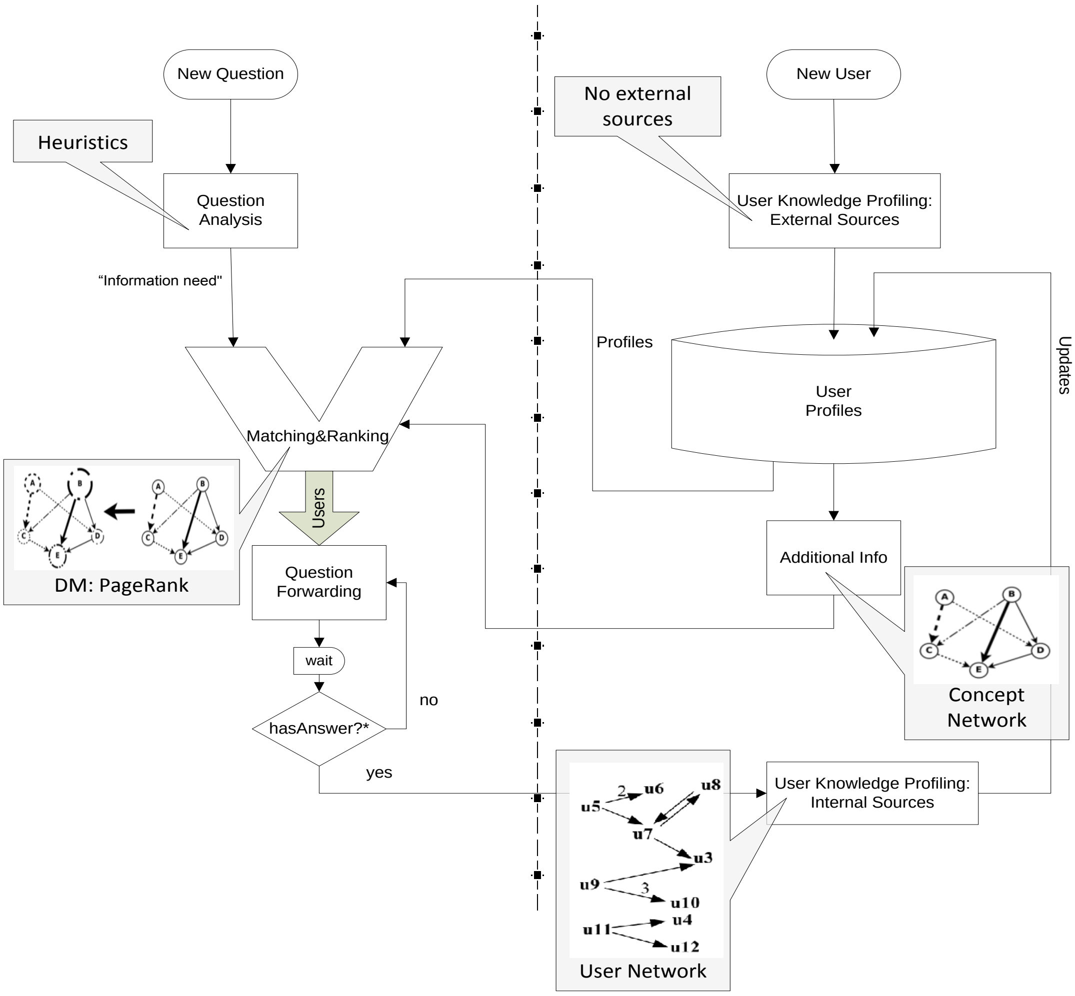

Figure 7. The G-Finder: Structure

answers. Therefore, algorithm makes correct guesses that they are likely to give the answers but unable to distinguish between real answers and commentary texts.

F. Aardvark

Horowitz D. and Kamvar S. D.[13] in 2010 presented a commercial system named Aardvark. It represents a social search engine where users can ask a question, either by instant message, email, web input, or voice. Questions are analyzed with DM technique (a combination of trained topic classifiers), and user can additionally annotate them with tags. To find someone that is most likely to be able to answer a question, Aardvark routes this question to persons in the user's extended social network. Therefore, user knowledge profile incorporates extended social network, which indexes affiliation and friendship information for every user and their friends, representing a Friends-of-Friends social graph. User has an option of importing this information from existing social networks like Facebook, LinkedIn, or webmail contacts, or manually inviting friends to join. Simultaneously, for *User Knowledge Profiling* from text, Aardvark maintains the list of topics about which the user has some level of interest. These topics are identified from several sources: manually indicated by user or a friend that invites him/her, parsed from a profile page or an account on which he/she regularly posts status updates (e.g., Twitter or Facebook) and finally, observing the user's behavior on answering (or electing not to answer) questions about particular topics. A combination of DM & NLP is used for topic extraction, particularly support vector machine & ad-hoc named entity extractor. Also, attributes like Connectedness and Availability are maintained for every user as *Additional Info*. Both extended social network and topics extracted from text are used to match & rank potential answerers. Similarity between users is computed using recommender systems techniques over extended social network attributes, like demographic similarity, profile similarity, social connections, etc. Similarity between extracted topics from question and topics from user's profile is calculated using corpus-based semantic similarity, which is computed over Wikipedia and other text corpora. Aardvark model organization is centralized and its structure is illustrated in Figure 8.

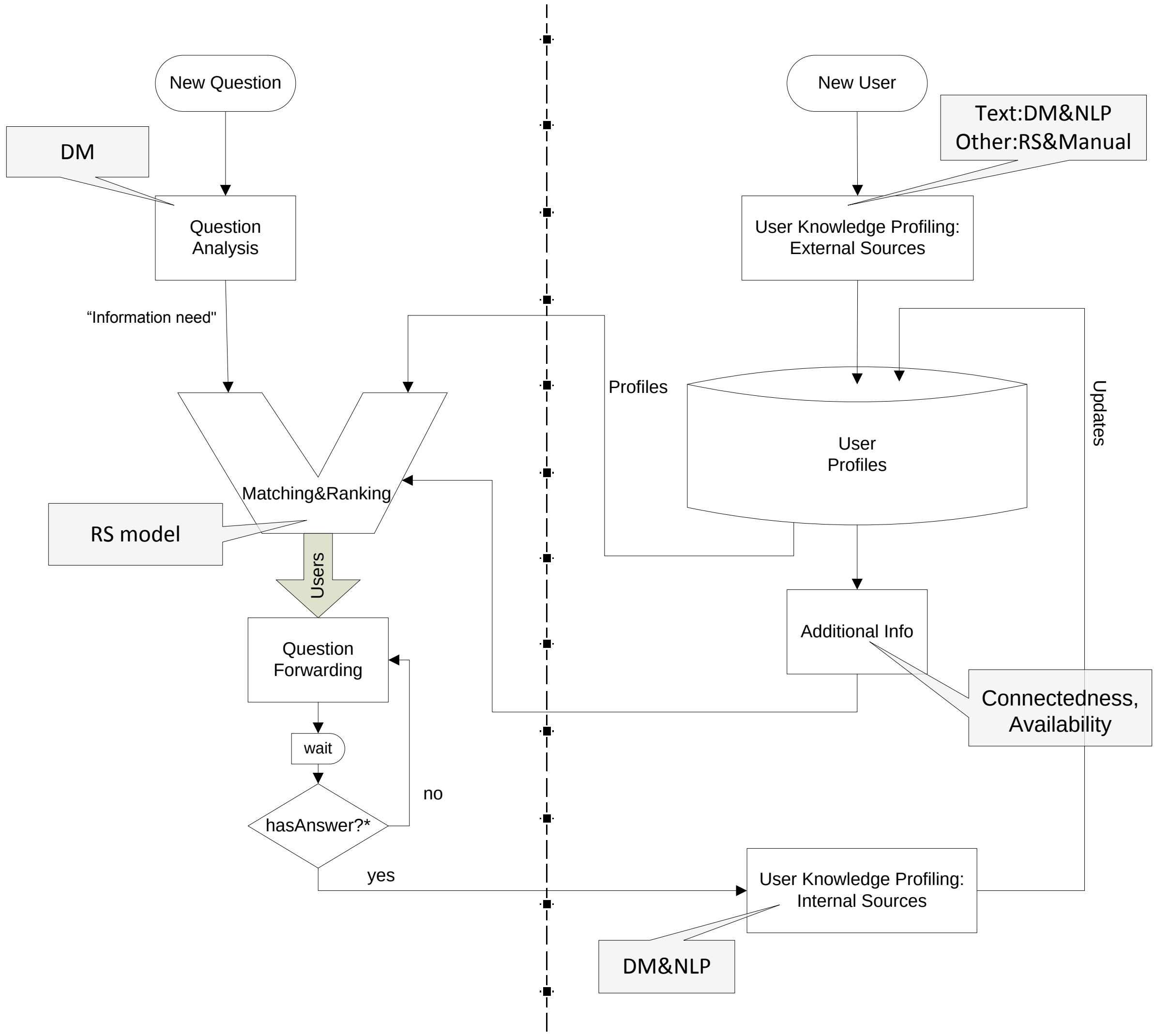

Figure 8. The Aardvark: Structure

The Aardvark search algorithm is being put on intimacy, where the user's trust in received answer is based on knowing the answerer (directly or indirectly from the social network). Thus, questions are routed primarily within user's extended social network. As reported, this provides results for questions that are in a context of user's social or demographic proximity (e.g., giving opinion about restaurant nearby or advice about dating). However, there is another dimension of trust in the received answer, based on the answerer's knowledge reputation. This particularly stands for questions that are highly expert-oriented, where the user's information need possibly cannot be satisfied within its social network. In this context different user's knowledge profiling model is needed in order to forward questions to competent users.

G. Confucius

Si X. et al[14] in 2010 introduced Confucius, a Google Questions&Answers (Q&A) service that employs six data mining subroutines to exploit synergy between web search and social networks. When search terms indicate a wh-query (when, where, why, etc.) or when the search engine cannot return sufficiently relevant results (e.g., the content overlap between query and potential top result pages is low), a Q&A session is recommended, encouraging the searcher to post a question on Confucius. Parallel implementation of Latent Dirichlet Allocation (PLDA) is employed for *Question Analysis*. Once a question has been typed in, a set of category labels are suggested for selection. Also, new labels can be manually added. While user is typing the question and before sending it to a human answerer, a question recommendation subsystem searches for similar earlier questions and their already available answers. This step reduces the time that takes for the user to obtain a satisfactory answer if a similar question exists. In addition, after submitting the question, complex and time-consuming NLP techniques are employed to automatically generate answers. This step does not compromise the overall system quality, since it is carried while waiting for the human-provided answers and if the confidence level of generated answer is lower than a predetermined threshold, it will not return any answers.

The *User Knowledge Profiling* module incorporates several features like the graph of users, the quality of answers, and the topic of interactions. For each answer, besides gathering askers feedback (in the form of best-answer selection or agree/disagree votes), an automatic answer-quality assessment routine is employed to produce quality ratings. This routine is implemented as a trained binary classifier and it is based on several factors including answer relevance to the question and its originality. Quality scores are then aggregated by *Matching&Ranking* routine, which quantifies user contributions and ranks users in a topic-dependent manner. This routine employs a weighted and topic-sensitive HITS computation over user activity graph, which is built by converting the Q&A threads to topically weighted interactions between users. The result of HITS captures two aspects of a reputation: (a) a user's ability to reach out to the other users and (b) a user's ability to gain attention from the others. The final score is then calculated as a linear combination these two. Also, the semantic matching is implicitly employed through the PLDA model. Confucius system is centrally organized. The structure is illustrated in Figure 9.

Reported statistics and evaluation results of the Confucius system showed that synergy between web search and community Q&A improved service quality. Also, to avoid redundant work (and unnecessary delay) by sending questions directly to human users, many techniques are employed, such as NLP generated answers or recommending similar and already answered questions. Nevertheless, the problem of Opinion Questions is one of the remaining challenges. The model training relies on best answers as positive samples and non-best answers as negative samples. However, for opinion questions, the distinction between a best and non-best answers is

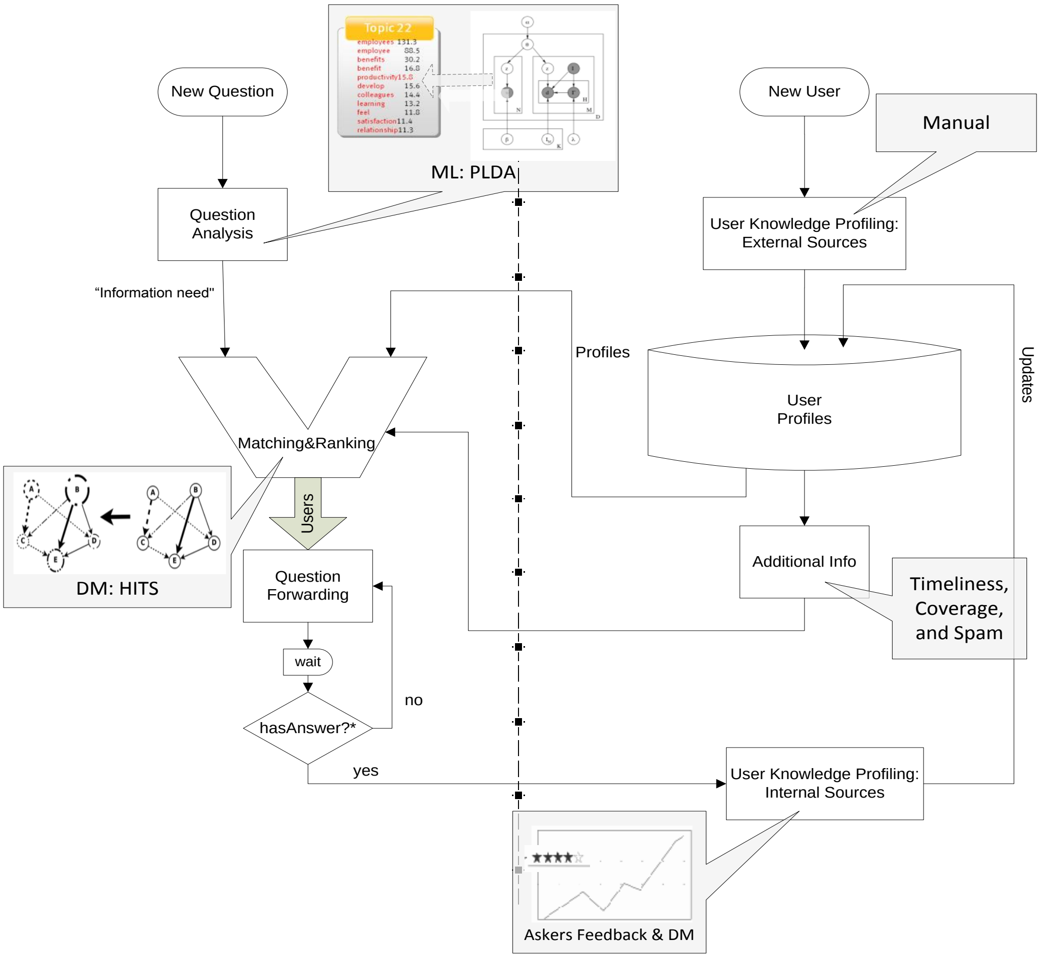

Figure 9. The Confucius: Structure

subjective. To handle this adequately, the model needs to capture the asker's subjectivity (like in Aardvark), which is not possible with the current set of features.

H. Yahoo! Answers Recommender System

Dror G. et al[15] in 2011 addressed a need for a mechanism in CQA portals (Yahoo! Answers in particular) that would expose users to questions they can relate to and possibly answer. *Question Analysis* is implemented using NLP techniques: stemming, stop word removals, and POS processing and filtering. Also, user has to annotate questions by assigning them to categories. System's architecture is centralized and *Matching&Ranking* is based on the multi-channel recommender system technology. To fuse and generalize information that represents multiple social and content signals from users, a single symmetric framework is constructed which incorporates and organizes these signals according to channels. Content signals are used for *User Knowledge Profiling* from text and they relate mostly to text attributes and categories of questions and associated answers. Other attributes are also included in a form of social signals, which capture the various user interactions with questions, such as asking, answering, voting, etc. The structure of the system is illustrated in Figure 10.

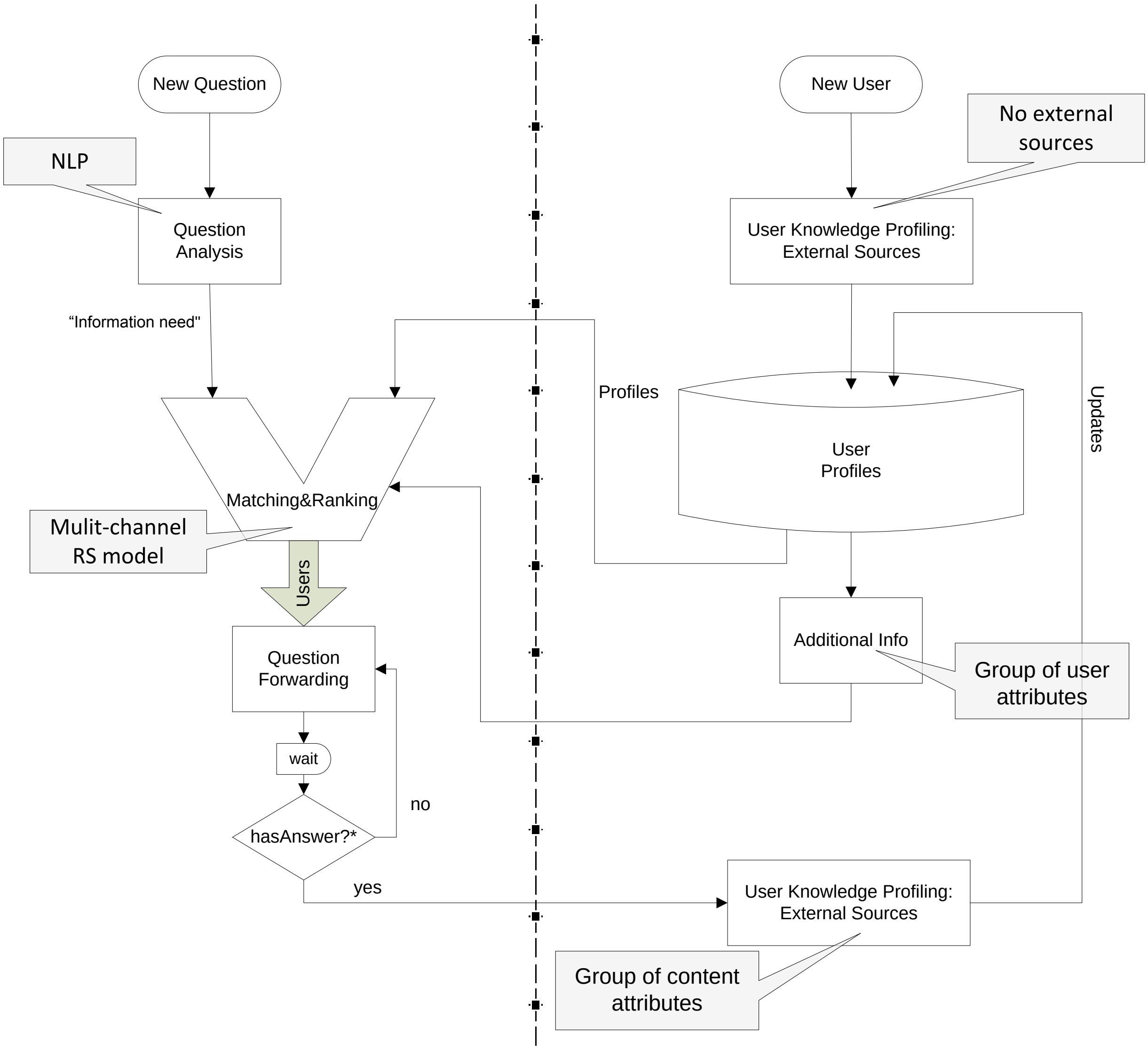

Figure 10. The Yahoo! Answers Recommender System: Structure

The key objective of this approach was to satisfy the sole asker in a variety of questions, some factoid but many being subjective where the notion of expertise is irrelevant. This differs from expert search task that tries to identify an authoritative answer that would satisfy most. Also, in a context of Yahoo! Answers system external sources of social relations between users are not available, so the main focus was to differentiate between various user-question interactions. This implicates the new user problem.

I. STM in CQA

Riahi F. et al[16] in 2012 investigated the suitability of two statistical topic models for solving an IQRS task and compared these methods against more traditional Information Retrieval approaches. The main objective of their study was to compute the probability that captures the expertise of each user for a posed question. In their model, each question has three parts: question tag – a tag assigned by user who posted the question, question title

– a short description, and question body - a detailed description. The interests of users are modeled by tracking their answering history in the CQA community. For each user, a profile is created by combining those questions answered by the user for which he or she has been selected as the best answerer. Based on the user profiles, the similarity between the answerer and a new question (*Matching&Ranking*) is measured by using a number of different methods such as TF-IDF, language models with Dirichlet smoothing, the Latent Dirichlet Allocation (LDA) and Segmented Topic Model (STM). Stemming, the stop-words removal, and the less-frequent-words removal are employed for questions and answers processing. General organization of the model is centralized and its structure is illustrated in Figure 11.

To evaluate different methods authors constructed a dataset from the Stackoverflow website. Results on this dataset showed that employed topic models outperformed other word-based methods. Further, authors concluded that STM gives consistently better performance compared to LDA. However, an advantage of external information sources for modeling expertise and interest of a user was not considered, as well as other internal attributes like answers score, favourite count and last edit date. Also, as authors note, user information often

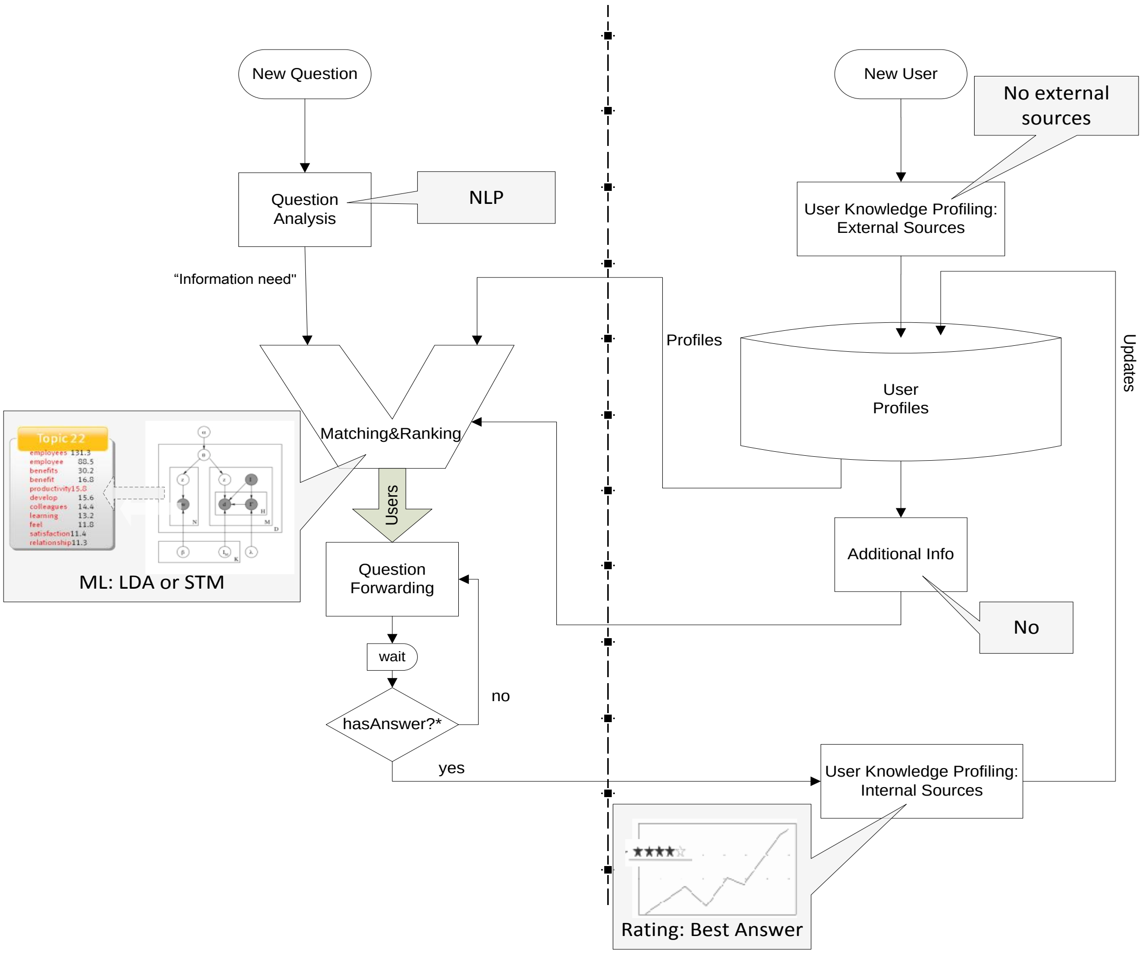

Figure 11. The STM in CQA: Structure

contains metadata information such as badges and reputation, which were not currently observed as additional variables (*Additional Info*) for profiling users.

J. Classification-based Routing in CQA

Zhou T.C. et al[17] in 2012 treated the IQRS problem as a classification task, and developed a variety of local and global features which capture different aspects of questions, users, and their relations. Local features include (a) question features (e.g., title length, body length, 5W1H question type - why, what, etc.), (b) user history features (e.g., member since, total points, number of best answers provided, number of questions asked, etc.), and (c) question-user relationship features (e.g., is a user the top contributor in the category the question belongs to). In addition, some features are adopted which describe the semantic similarity of the question's language model and the user's language model (e.g., the KL-divergence between the current question's title/detail and all the questions' title/detail the user has provided answers). Features referred as global take into account the global information of the CQA service about a user, which can also be divided into the same three categories: (a) question features (e.g., average title length and average detail length), (b) user history features that capture uniqueness of the user (e.g., KL-divergence of the user's answered, asked, and starred questions with all users' answered, asked, and starred questions), and (c) question-user relationship features, which are based on an assumption that the more similar are language models of user's answered questions and the questions in a category, the more probable a user would answer the questions from the category (e.g., KL-divergence value of a user's answered questions' title/detail with questions' title/detail in the category the given question belongs to). Stemming and the stop-words removal (except 5W1H words) are employed for questions and answers processing. The *Matching&Routing* task is treated as a two-class classification problem, but the focus was on the positive class, which means given a question-user pair, would the user answer the question. Support vector machines (SVM) was employed as the classification method. Model organization is centralized and its structure is illustrated in Figure 12.

Experimental results were obtained from an evaluation over the Yahoo! Answers dataset. If a user answered a question, the question-user pair is considered as a positive instance, and if a user asked a question, the question-user pair is considered as a negative instance. The reason for the last is that if a user asked a question, it might mean that he/she did not possess the knowledge about the question, and this indicates the user was unable to answer the question. But as authors of the paper admit, the procedure of choosing negative instances is arguable. Also, a comparison was performed on how different types of features contribute to the final results. This showed that question-user relationship features play a key role in improving the overall performance.

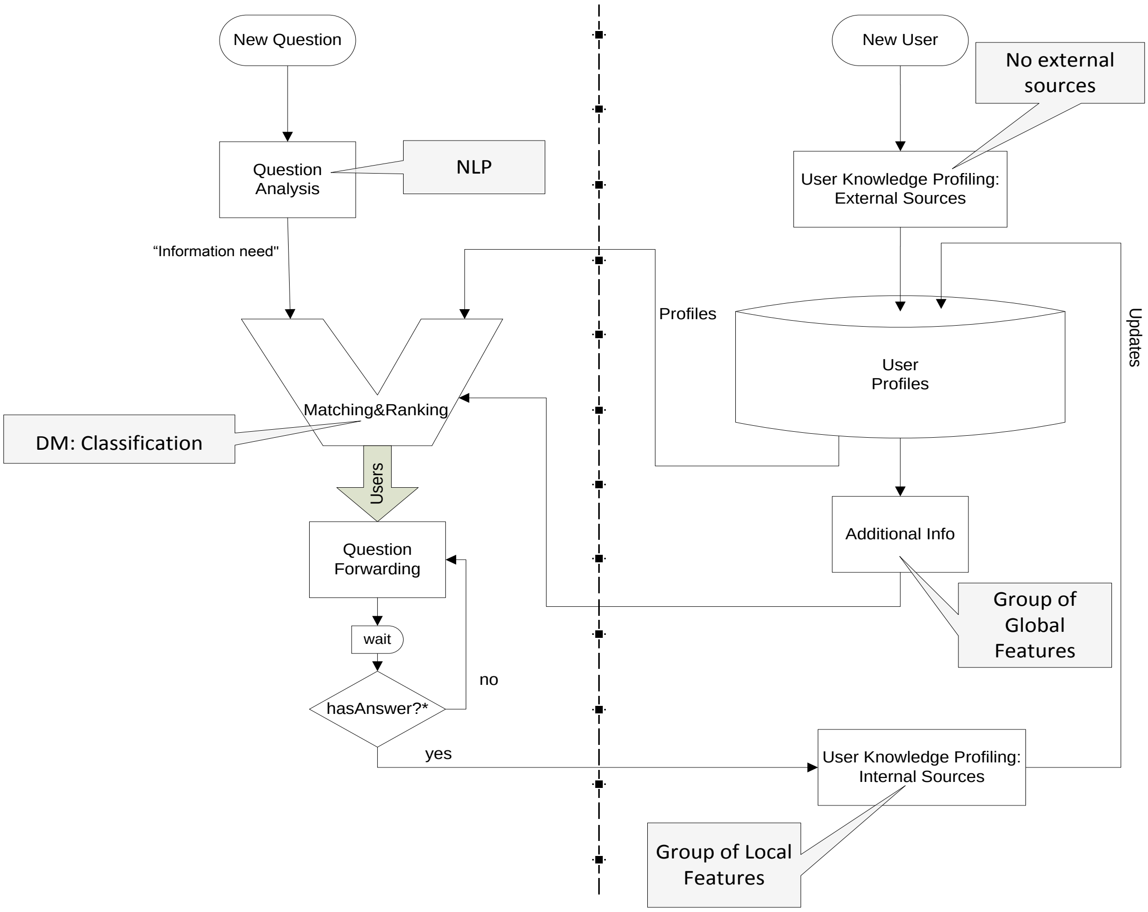

Figure 12. The Classification-based Routing in CQA: Structure

However, the new user problem exists as well as the problem of opinion questions, in which the distinction between a best and non-best answers is subjective.

K. SQM

Banerjee A. and Basu S.[18] in 2008 proposed Social Query Model (SQM) for decentralized search, which has the Pagerank model and certain Markov Decision Processes as special cases. The social network is represented as a graph with nodes and links. The model does not consider question analysis and there is no question annotation. The organization is decentralized and Matching&Ranking is based on a distributed approximation algorithm, which computes optimal query routing policy. User's knowledge profile includes only expertise score and as an Additional Info response rate is incorporated. Therefore, in the context of the model this policy is simultaneously optimal for all nodes, in that no subset of nodes will jointly have any incentive to use a different local routing policy. Illustration of the SQM model structure is presented in Figure 13.

To some extent, all previously presented approaches are complementary to SQM, since the focus was not on query routing within the nodes of a social network, but on identifying a user's potential to give the correct answer

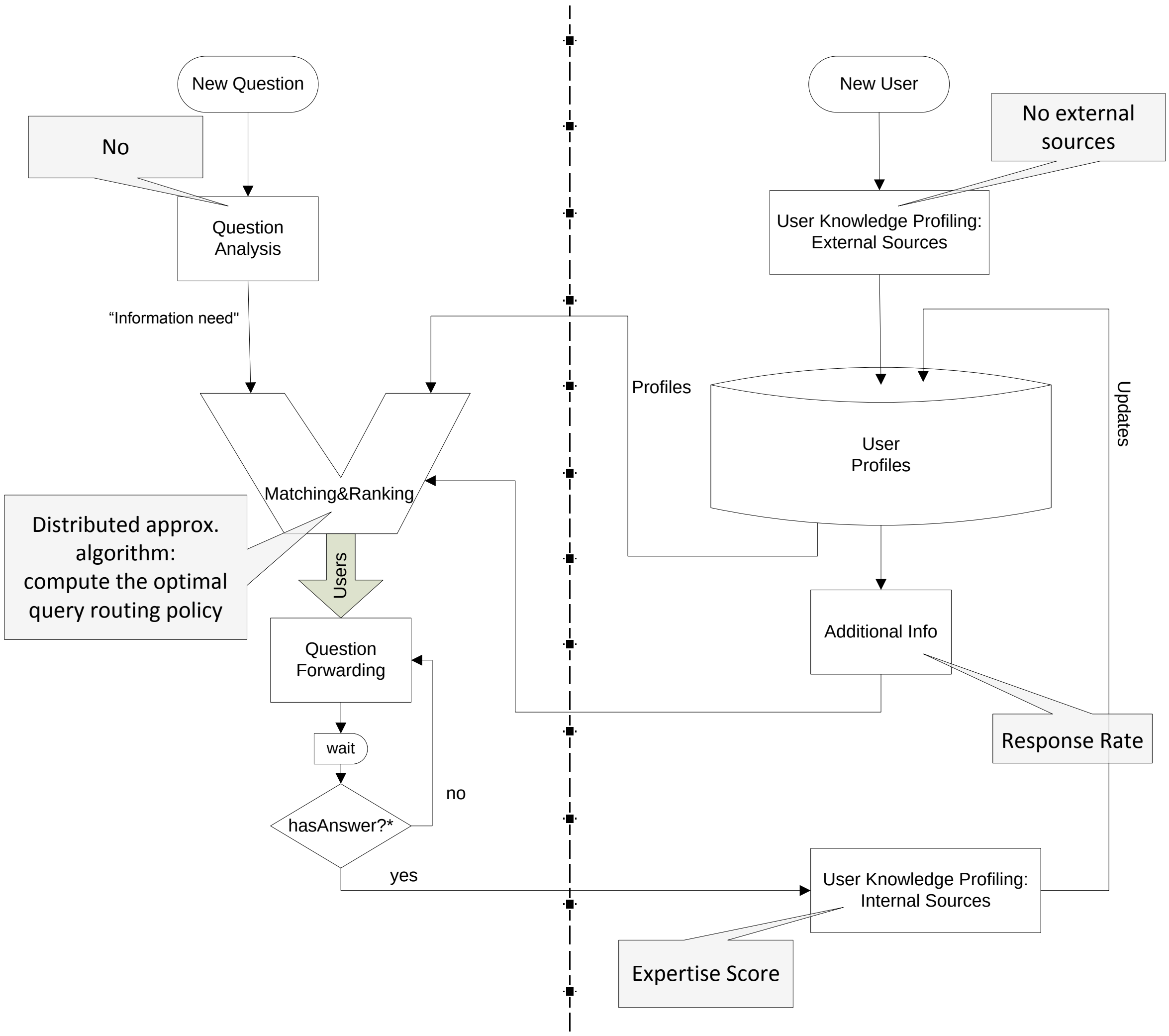

Figure 13. The SQM: Structure

and matching that potential to particular question. Therefore, the potential can be characterized by different cfactors, such as expertise and responsiveness, which are input parameters within SQM.

V. EVALUATION AND FUTURE RESEARCH

Based on the analysis of previously presented solutions, the following issues were identified and the rest of this section contains ideas of potential interest for future research related to those issues.

A. Question visualization

Questions typically consist of a text, which is not too long, so one solution is that a *Question Analysis* module can be developed using NLP or DM/ML techniques. However, tools for automatic information extraction, in general, can be insufficiently precise and can omit some valuable information. Also, short questions are often ambiguous. Having that in mind, among all analyzed, the best proposed solution is Confucius, since for a typed question a set of category labels are suggested for selection. Also, new labels can be manually added.

Consequently, the most effective solution is an interactive user interface that allows communication between the *Question Analysis* module and a user that posed the question. This approach combines fully automatic text processing and manual correction of results, giving the user a possibility to increase the accuracy of the output. On the other hand, automated processing can produce more results that would be usually forgotten.

Above that, we propose that the discovered concepts should be visually represented in a form of a concept cloud (TagCloud visualization). One benefit of this approach is that "the more significant the concept is, the bigger the font size it has," which provides a more intuitive representation of relations between specific concepts and their importance in the question. Figure 14 represents an example of generated concept cloud. More details about possible implementation can be found in [19].

B. Semantic and string similarity incorporation

The measure of similarity between a question and a user profile can be realized by computing exact mach between recognized topics or terms, or by calculating their semantic similarity. As mentioned, posted questions are usually short in length, so the question can be semantically similar to the user profile but still lexically very different. Therefore, better results can be achieved with a system that is capable of capturing the semantic similarities. Several solutions like Aardvark, Confucius, and G-Finder use semantic similarity within the *Matching&Ranking* phase. However, questions or profiles can include typos or different forms of infrequent proper nouns which can decrease performances of the system.

One possible solution is the use of a bag of words approach, which can employ corpus-based or knowledge-based measures of word similarity[20]. For each word in a profile, the method should identify the highest match from the question and then combine it in the overall measure of semantic similarity. Islam and Inkpen[21] proposed an improvement of this similarity measure by incorporating a string matching algorithm with a corpus-based measure of semantic word similarity. Therefore, we indicate that a possible improvement can be found in this direction as illustrated in Figure 15. This method, besides the semantic word similarity measure,

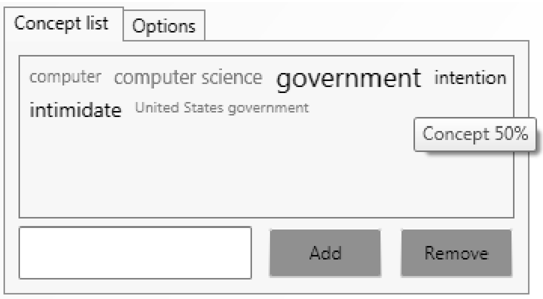

Figure 14. Question Visualization: An Example of a Generated Concept Cloud

incorporates the string similarity measure, so it performs better with typos, evolving hotwords or different forms of infrequent proper nouns[22].

C. Profile Integration

Issues related to profiling of user knowledge are: (1) utilization of Bayesian probability, (2) the new user problem, and (3) the integration problem. (1) Bayesian probability has a firm theoretical foundation and it is widely used in trust management, at present. However, the Bayesian approach does not have an adequate expressiveness and it needs some artificial construction. For example, user A answered 100 questions about a topic *c* and the quality of the answers rated by other users was 0.5. Then, we consider another case that A did not answer any question about topic *c* (there is no information about A's knowledge in topic *c*). In both cases the evaluated trust of the Bayesian approach in A's knowledge about the topic *c* is p(trust)=0.5 and p(distrust)=0.5. Therefore, the Bayesian approach does not have the ability to distinguish these two cases[23]. (2) The new user problem can be also found in the Recommender Systems domain of research[24]**,** since it is difficult to make a profile for a new registered user. Approaches like Aardvark or iLink use external sources (e.g., information from social networks, blogs, or manual input) to make initial user knowledge profiling, and also for following updates. (3) This kind of approach introduces the third problem, the integration problem, which relates to an issue how to seamlessly integrate information about the user from different sources. Therefore, a new approach for user knowledge profiling is needed, which will enable uniform incorporation of various information sources in the form of software agents.

One possible direction toward solution can be found in the Dempster-Shafer theory (DST), a mathematical theory of evidence that is a generalization of the Bayesian probability. It can handle ignorance naturally and allows one to combine evidence from different sources. To arrive at a degree of belief (represented by a belief function) DST takes into account all available evidences. In addition, for profile integration within IQRS we

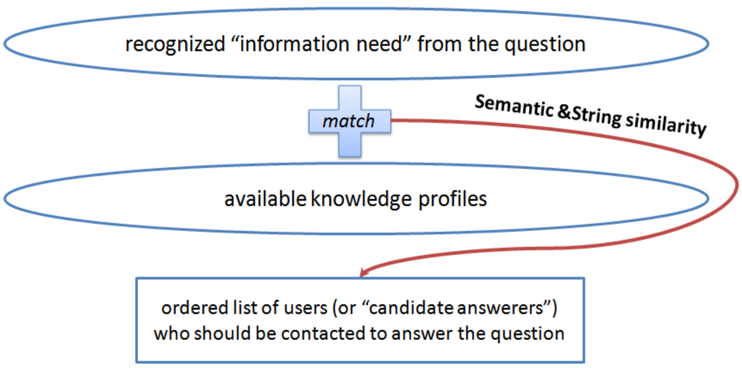

Figure 15. Incorporation of Semantic and String Similarity: Ilustration of Matching Process

suggest the use of an evidential trust model based on the Dezert-Smarandache theory[23], which is a generalization of DST, so it has a higher expressiveness.

VI. Conclusion

The objective of this paper was to systematically establish common characteristics of IQRSs, which are inherently heterogeneous, and to allow their uniform analysis. Hence, major contributions of this survey paper are: identification of the IQRS domain of research, generalization of approaches and the original presentation paradigm, description of analyzed approaches and their comparative study, and finally a proposal of three new directions for future research, which are related to the three main issues of importance for intelligent question routing. To the best of our knowledge, this survey includes all approaches related to IQRS published in the open literature after 2007.

Findings and explanations of this survey are of interest to those who like to enter this emerging field of research, to understand the essential notions, and to obtain ideas for their future research. This may be of most benefit to PhD students.

Newly open problems fall into two basic categories: (a) to expand the survey to a wider body of knowledge and (b) to implement a prototype of the proposed new ideas and to evaluate their performance, comparatively with the best solutions from the open literature, or their approximated equivalents.

In conclusion, since questions and appropriate answers are the essence of IQRS, our attitude in this paper was: "*Prudens quaestio dimidium scientiae* - Half of science is asking the right questions" Aristotle (384 BC – 322 BC). Or more appropriate: “Half of answer is the right question” Therefore, we asked the three fundamental questions and on their basis we built the presentation paradigm, which is supposed to be the main contribution of this paper.

Acknowledgment

Authors would like to thank Prof. Nenad Mitic for his advices and constructive discussions. The work presented here was partially supported by the Serbian Ministry of Education and Science (projects III 44006, III 44009, and III 32047).

References

1. Dang HT, Lin J, Kelly D, Hill C. Overview of the TREC 2006 Question Answering Track. In: Proc TREC, 2006.
2. Ounis I, Macdonald C, Soboroff I. Overview of the TREC-2008 Blog Track. In: Proc TREC, 2008.
3. Lappas T, Liu K, Terzi E. A survey of algorithms and systems for expert location in social networks. In: Aggarwal C, editor. Social Network Data Analytics: Springer; 2011. p 215-241.
4. Faye D, Nachouki G, Valduriez P. Semantic Query Routing in SenPeer , a P2P Data Management System. In: Proc Network-Based Information Systems; 2007. p 365-374.
5. Tempich C, Karlsruhe D, Staab S, Wranik A. REMINDIN : Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors. In: Proc WWW, 2004.
6. Rus I, Lindvall M. Knowledge management in software engineering. IEEE Software 2002;19:26-38.
7. Frameworks S. Management of explicit and tacit knowledge. Journal of the Royal Society of Medicine 2001;94:6-9.
8. Davitz J, Yu J, Basu S, Gutelius D, A. iLink: search and routing in social networks. In: Proc WWW, 2007.
9. Qu M, Qiu G, He X, Zhang C. Probabilistic question recommendation for question answering communities. In: Proc WWW, 2009. pp 1229-1230.
10. Zhou Y, Cong G, Cui B, Jensen CS, Yao J. Routing Questions to the Right Users in Online Communities. In: Proc ICDE, 2009. pp 700-711.
11. Li B, King I. Routing questions to appropriate answerers in community question answering services. In: Proc CIKM, 2010. pp 1585-1588.
12. Li W, Zhang C, Hu S. G-Finder: Routing Programming Questions Closer to the Experts. In: Proc OOPSLA/SPLASH, 2010. pp 62-73.
13. Horowitz D, Kamvar SD. The anatomy of a large-scale social search engine. In: Proc WWW, 2010. pp 431-441.
14. Si X, Chang E, Gyongyi Z, Sun M. Confucius and its intelligent disciples: integrating social with search. In: Proc VLDB, 2010. pp 1505-1517.
15. Dror G, Koren Y, Maarek Y, Szpektor I. I want to answer; who has a question?: Yahoo! answers recommender system. In: Proc KDD, 2011. pp 1109-1117.
16. Riahi F, Zolaktaf Z, Shafiei M, Milios E. Finding expert users in community question answering. In: Proc WWW – CQA Workshop, 2012. pp 791-798.
17. Zhou TC, Lyu MR, King I. A classification-based approach to question routing in community question answering. In: Proc WWW – CQA Workshop, 2012. pp 738-790.
18. Banerjee A, Basu S. A social query model for decentralized search. In: Proc SNA-KDD, 2008.
19. Varga E, Furlan B, Jakus G, Milutinovic V. Document filter based on extracted concepts. Transactions on Internet Research 2010;6(1):5-9
20. Mihalcea R, Corley C, Strapparava C. Corpus-based and knowledge-based measures of text semantic similarity. In: Proc National Conference on Artificial Intelligence, 2006. pp 775-780.
21. Islam A, Inkpen D. Semantic text similarity using corpus-based word similarity and string similarity. ACM Transactions on Knowledge Discovery from Data 2008;2:1-25.
22. Furlan B, Sivački V, Jovanović D, Nikolić B. Comparable Evaluation of Contemporary Corpus-Based and Knowledge-Based Semantic Similarity Measures of Short Texts. Journal of Information Technology and Applications 2011;1:65-71.
23. Wang J, Sun H-J. A new evidential trust model for open communities. Computer Standards & Interfaces 2009;31:994-1001.
24. Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions On Knowledge And Data Engineering 2005;17:734-749.