An Approach on Improving Search Engines through Social Content Recommendation

Una aproximación a la mejora de los Motores de Búsqueda a través de la Recomendación Social de Contenido

Luis Alberto Pérez García
Universidad Europea de Madrid / Universidad Rey Juan Carlos de Madrid
--- dirección postal --
perezgarcia.luisalberto@gmail.com

Resumen: El crecimiento de Internet ha provocado que la búsqueda de información haya pasado a tener uno de los papeles más relevantes de la industria y a ser uno de los temas de mayor actualidad en los ambientes de investigación. La red de redes es el mayor contenedor de información de la historia y su facilidad para generar información conlleva nuevos retos a la hora de recuperar dicha información y discernir aquella que tiene mayor relevancia que el resto.

Paralelamente al crecimiento de la información en cantidad, también ha cambiado la forma en que podemos acceder a dicha información. Uno de los cambios que más movimiento de información ha provocado ha sido la aparición de las **redes sociales**. Hemos podido ver como las redes sociales pueden llegar a provocar más tráfico de información que los propios buscadores. Indudablemente, podemos sacar algunas conclusiones que nos permitan dar un enfoque ligeramente distinto al problema de la recuperación de información: el público general confía más en el contenido que le llega a través de contactos conocidos.

En éste documento exploraremos un posible cambio en los motores de búsqueda clásicos para hacerlos más sociales.
Palabras clave: búsqueda, motores de búsqueda, búsqueda social, redes sociales, recomendación de contenido, recuperación de información, sistemas de recomendación de contenido

Abstract: Internet growth has provoked that information search had come to have one of the most relevant roles in the industry and to be one of the most current topics in research environments. Internet is the largest information container in history and its facility to generate new information leads to new challenges when talking about retrieving information and discern which one is more relevant than the rest.

Parallel to the information growth in quantity, the way information is provided has also changed. One of these changes that has provoked more information traffic has been the emergence of **social networks**. We have seen how social networks can provoke more traffic than search engines themselves. We can draw conclusions that allow us to take a new approach to the information retrieval problem. Public trusts the most information coming from known contacts.

In this document we will explore a possible change in classic search engines to bring them closer to the social side and adquire those social advantages.
Keywords: search, search engines, social search, social netwoks, content recommendation, information retrieval, recsys

1 *Introduction*

1.1 Search so far

1.1.1 Classic Search

In classic information retrieval (IR) systems, the main goal is to return high-relevance and high-quality information.

Precision is an important term in IR systems. It means the percentage of documents retrieved that are revelant for the query.

An IR system needs to keep some process sovered to be functional, some of them are quite common: gather existant information, convert that information to an structured model, ponderate that documents according to its relevance and retrieve documents according to a query.

Although all this processes are necssary, the one which ponderates documents is nowadays the most important and the one which makes a difference between search engines.

1.1.2 Measuring the Relevance

The way relevance is measured has always been related to the topic of the document. In classic search retrieving the document which was the most accurated to the query was good enogh. With the growth of information, we can find lots of documents that really tight to each topic from long ago. We needed a method for discerning which of those documents were the best result.

This was the time when Google comes to scene. They provided a revolutionary algorithm to measure relevance: the PageRank.

Google's pagerank takes care of which documents are the most selected by users searching about a same topic. It is autoadaptative, and best results are shown in best postitions for users to find them easily.

This algorithm brought a revolution to information retrieval and has been the best solution so far.

1.1.3 Problems

Focusing no the pagrank, we can see that it works in only one way, clicked results get a favorable treatment being promoted to higher positions while not clicked results remain in lower positions in the result page. If we extend this behavior for a long time it is easy to understand that the best old documents will continue being treated as the best ones even if a better new document appears.

It also brings to the table a new problem, every clicked result is a good one? Obviously not. Maybe the title for a document is good enough to give it a click but when reading it we notice it is not worth the position it has. The necessity of a recommendation system that allows to evaluate the documents **after** seing them is a priority.

Another problem is that measuring the success of a search based on the results relevance according the query can be a mistake. Who do we really want to satify? The query introduced or the user who introduces the query? Obviously we want the user to find what he is looking for, even if the query is not the best approach. Google has made some efforts on this, using query suggestion, but it is not good enough as this approach is global oriented. A social approach could result in a better solution here.

1.2 Impact of social networks

1.2.1 New Kings of Information

The way the information reaches users is changing. Some time ago there were two main ways to get to the information: direct visitis to kown sites and results from search engines

Lately we have witnessed a change here. Information is reaching users from social networks. This is due to people finding more relevant information from their contacts than from search engines. This year, Facebook overtake Google in traffic, also Twitter is driving more traffic each day.

If this trend keeps going, we could be talking about a change of reign in information retrieval quite soon.

1.2.2 Personalized Model

The great advantage of social networks is that you only see the information you want to see, and you can filter it by selecting which contacts are related to you.

Social networks can be seen as a new way of retrieving information where the user decides who are the sources that provide his information.

Facebook can be the better example of a classic social network. It allows you to keep track of your contacts and see information they are producing or information they find important. That is the great point, they keep in mind that possibly you are interested in the same information your contacts are.

Twitter is, so far, the best mixture of search engine and social network. It is oriented to information and its main point is being the fastest spreading news. It keeps the advantage of being social, as the information user gets from Twitter is provided by its social contacts and uses a, still quite poor, topic filtgering system to help discovering the best information providers.

1.2.3 Advantages

There are tow huge components in getting information: the information you retrieve when searching for it and the information that reaches you altough you are not looking for it.

In a classic environment, the information that reaches you directly is usually treated as *unwanted information* because it barely results to be relevant to you.

The social component is making a point here. Most of the information we consume, without considering its relevance, comes directly to us. A social approach can make that information to be relevant and that fact is an advantage that search engines can not let go through.

2   State of the Art

2.1 Social Search

2.1.1   Existing Approaches

Social Search is quite an unexplored field. Some approaches has been made, such as the *Village Paradigm* and some implementations of these approaches are working fine, such as Aardvark.

The village paradigm is based on finding the right person to answer a query, rather than the document that contains the information that answers the query.

Using the village paradigm remains more to a recommendation system than to a search engine. In fact, what the village paradigm does

is recommending you the person that can answer your query the best.

An important point here is that you need people wondering to answer questions. Seen Aardvark results, people do like to answer questions but, remaining your social search engine base on users can be a failure if the number of active users falls down.

It is to be noticed that the village paradigm works very well on queries that deal with opinion, advice, experience or recommendations. When having a conventional query that is factual or navigational, and it is based on keywords, the *library paradigm*, which is the one used in classic search engines, performs better than the village one.

Working with the village paradigm makes necessary to have an statistical model for routing questions to potential answerers and also needs a method for indexing people. Both things are relationated, users are indexed according to the topics where they knowledge is relevant, and queries are routed to people having a relevant knowledge on the query topic.

An extra problem to deal with when working on social search is understanding the natural language. Queries in social search engines based on the village paradigm tend to be questions in natural language, usually related to locations, times and more specific issues. It is needed to analyze the query and extract relevant information to route it according to its topic. It is not as hard as understanding natural language per se, but the necessity of understanding the topic of a query is one of the important steps.

3   System Description

3.1 Concept and Goals

3.1.1   Main Concept

The main goal of this document is to describe an approach to a mixture model of classic and social search. Both paradigms have been taken in count, the library and village one.

The idea is to make classic search engines more powerfull by adding some recommendation based con social realtionships. As Aardvark recommends answerers according to the introduced query this approach will deal with recommending the best piece of information according to a classic search and direct recommendations from users.

Some of the problems search engines are suffering where metioned before, and they can be mitigated using social recommendations.

3.1.2 Goals to Achieve

The main goal is to reponderate results from a search engines according to users recomendation and get a more relevant results organization.

Classic search engines do a reponderation on their own results based on the ones that get clicks from users. This makes that reponderation is done without taking in count if the results satisfy the user or not.

The approach proposed in this paper is similar to that one, but the reponderation will take in count the user satisfaction with the selected result, as recommendation or penalization will be done after the information is seen by the user.

Notice that this also allows to mitigate the SEO problem that affects to the most of classic search engines.

3.2 Architecture

3.2.1 Information Source

This initial approach will be held as a little experiment so, initally, there will be no self indexed information.

The information source of our social search system will be a classic search engine, specifically Google search engine.

This provokes that our social search system will not be able to work standalone, but will allow us to reduce the initial size of our database as we will only store information related to recommendations made by users.

3.2.2 Database

Information on recommendations gather by the system will be stored in a database as well as information related to users and their social relationships.

Common databases are not the best option for social based systems. Working with relationships between users makes the relational database model obsolete as it has a poor performance and lacks of scalability. The proper solution will be to use a graph database, which have been created with social needs in mind and are being adopted by the most important projects related to social technologies.

3.2.3 Ranking Algorithm

None of the existing algorithms, classic ones as the pagerank and the Aardvark one fits accurately on this model.

The algorithm needed will deal with two important factors: the selected result from the search engine and the recommendation given by the user.

An improvement of this algorithm will also include a user indexer to relate users and topics, as the Aardvark system does, so the system can give more relevance to reommendations made by users with a proven knowledge in the query topic.

This will also allow to apply the *"following/follower"* model used by Twitter, to search environments. Users could be notified when a relevant user recommends a result on a topic related to previous queries.

3.3 Usability

3.3.1 Simple Query Interface

One of the problems recommendation systems have to deal with is the user interface. Convencing the user to evaluate if something is recomendable or not can not be made throught a heavy process.

User interface to introduce a query should be minimalistic and clean. The result lists has to be simple but containing enough information to allow the user to select the result that fits best his query.

3.3.2 Recommendation

The interface for recommending a result should be the the least intrusive possible.

The idea is to show a small bar in the browser window which showing the content of the selected result. That bar will able the user to recommend the result and even to rate it.

4 Future Work and Conclusions

This is only a minimal specification on a different approach to social search so, many future works can be borned from here.

As a first option, a future work will be to replace the source search engine with a metasearch engine, which will retrieve unique results from different search engines.

A suggestion system for powerful users will be also a great addition to the system.

5   *Referencias bibliográficas*

El primer párrafo de cada sección no llevará sangría, los restantes sí.

The Anatomy of a Large-Scale Social Search Engine (Damon Horowitz and Sepandar D.Kamvar)

Information Retrieval – Introduction and Survey (Norbert Fuhr)

A Survey on Web Information Retrieval Technologies (Lan Huang)

Information Retrieval: A Survey (de Greengrass)