

Search on enterprise Wiki

Natalya Angapova

Yandex

Russia, Moscow

16, L. Tolstogo st.

+7(495)7397000

angapova@yandex-team.ru

ABSTRACT

In this poster, I share the results we have got while working on a better and easier search on our enterprise Wiki.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *search process*.

Keywords

Wikis, corporate, enterprise wikis, search.

1. INTRODUCTION

Yandex's corporate Wiki is about 50000 pages. It is 10% of russian Wikipedia. It consists of project documentation, todos, guidelines, code examples, etc. Wiki has 1300 visits and almost 1000-1200 search queries everyday. What is the difficulty to find something in corporate Wiki? First of all, there is no such redundancy as it is in Internet. A small number of documents, pages describes one concrete subject. Information is very separate and often it is not prepared to search on. Solution of the problem of relevant search on corporate Wiki is the major step to enterprise search in the company.

2. COLLABORATIVE SEO

The great difference between corporate wikis and Wikipedia or sites in Internet is nobody spends its time on search optimization. We have decided to urge our workmates to make wiki-pages more searchable. We have put the keywords in the body of the page. So it has become easy to specify and edit them for everybody. Next step is to give people human titles - titles of the pages they create - in snippets of the results on SERP (search engine page results). It will allow us to remove duplicate urls from snippets. We have analyzed every page and automatic changed wiki-URLs in snippets on titles we have found.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '10, July 7-9, 2010, Gdańsk, Poland
Copyright © 2010 ACM 978-1-4503-0056-8/10/07... \$10.00

Now we are going to put the titles out of wiki-pages in separate field, so everybody could edit them. As we think it will improve

search characteristics of pages. And maybe it's a first step to classification, taxonomy, of wiki-pages.

3. SERP

We think enterprise search should be fast and easy. SERP should be clear and helpful. That's why we develop correction misprints and suggest taking into account intranet-vocabulary of our company. We make navigational links in suggest and put them on the first places of SERP. We change the URLs in snippets on navigational chains. So wiki-structure becomes more visible and clear. And we are going to use quick-links in snippets for the purpose to make any understandable action quickly. For example, to sign on page changes. All these actions improve search capability in whole.

4. USER BEHAVIOR ANALYSIS

It is impossible to develop search without analyzing the behavior of users. For this we keep and analyze search logs: search queries, surfing on pages, clicks on search results. Also we have users assessments of results. They help us to teach and modify the ranking formula of search engine.

4.1 Search metrics

We have different search quality metrics which help us to understand what to do. We focus on them every time we update anything in search. Some of them:

1. Estimate of the likelihood that users will find information relevant intent query.
2. Average first click position.
3. Part of null results.
4. Part of results without clicks.

5. IMPROVEMENT

What else we have done that makes our search better:

1. Search with users access rights.
2. Make search in root wiki-cluster.
3. Search taking into account all possible page names.

What we are going to do:

4. Search taking into account synonyms and abbreviations.
5. Personalizing search.
6. Indexing attached files.
7. Realtime indexing.

8. Separation person and service clusters.

6.RANKING FACTORS

In addition to the above we also are going to experiment with search ranking factors. Some of them may have an effect, some of them not. What we want to try:

9. Page factors: update date, number of visits, bookmarks, number of watchers, etc.

10. Link ranking: incoming and outgoing links, link titles.

7.REFERENCES

There is no any reference in my work. Search engine concerned in the poster is our company's product. Everything is taken from our own experience, developments and experiments.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.