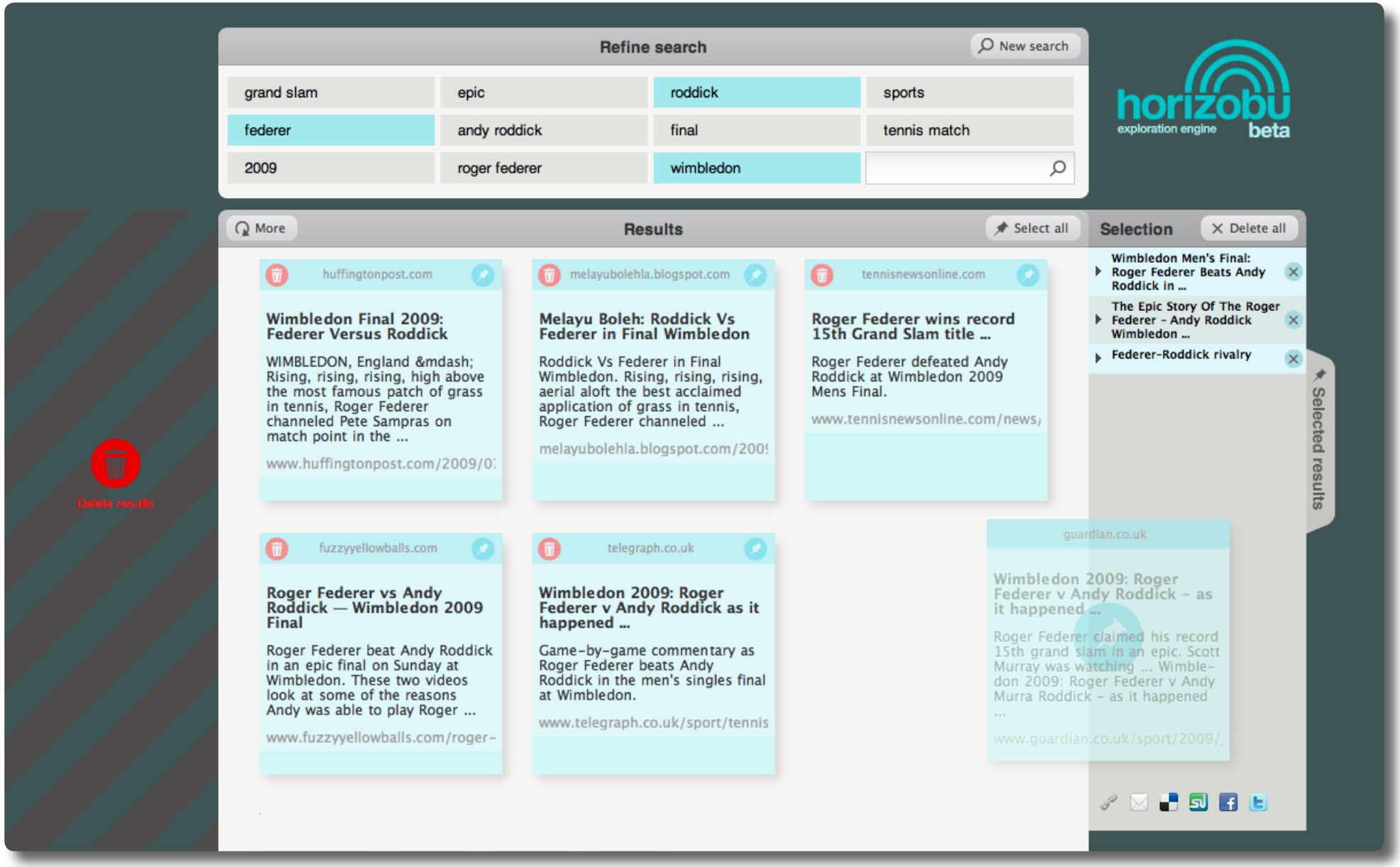
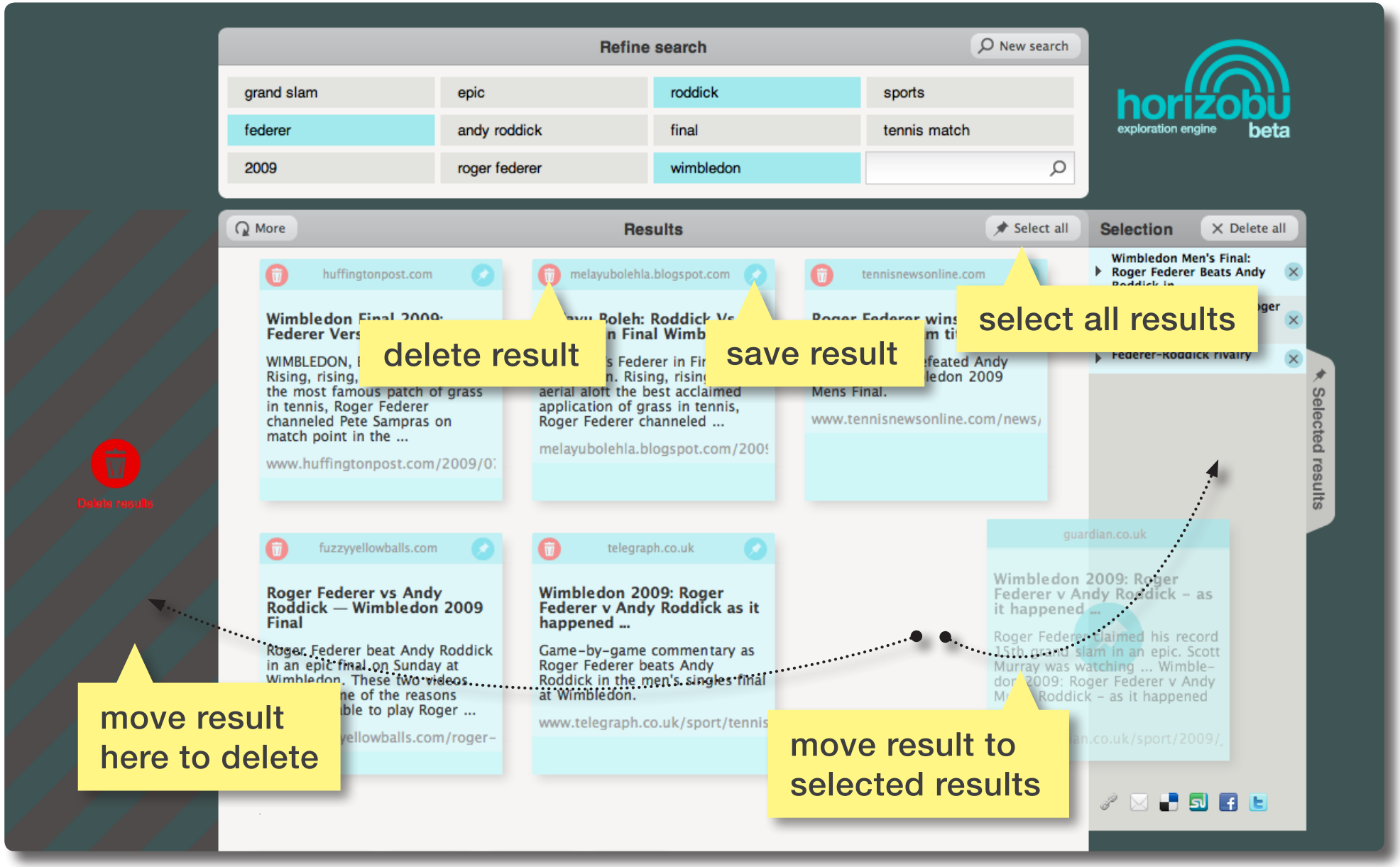

horizobu
exploration engine


Refine search
New search
grand slam
epic
roddick
sports
federer
andy roddick
final
tennis match
2009
roger federer
wimbledon
horizobu
exploration engine
beta
More
Results
Select all
Selection
Delete all
Delete results
huffingtonpost.com
Wimbledon Final 2009: Federer Versus Roddick
WIMBLEDON, England — Rising, rising, rising, high above the most famous patch of grass in tennis, Roger Federer channeled Pete Sampras on match point in the ...
www.huffingtonpost.com/2009/0
melayubolehla.blogspot.com
Melayu Boleh: Roddick Vs Federer in Final Wimbledon
Roddick Vs Federer in Final Wimbledon. Rising, rising, rising, aerial aloft the best acclaimed application of grass in tennis, Roger Federer channeled ...
melayubolehla.blogspot.com/200
tennisnewsonline.com
Roger Federer wins record 15th Grand Slam title ...
Roger Federer defeated Andy Roddick at Wimbledon 2009 Mens Final.
www.tennisnewsonline.com/news/
fuzzyyellowballs.com
Roger Federer vs Andy Roddick — Wimbledon 2009 Final
Roger Federer beat Andy Roddick in an epic final on Sunday at Wimbledon. These two videos look at some of the reasons Andy was able to play Roger ...
www.fuzzyyellowballs.com/roger-
telegraph.co.uk
Wimbledon 2009: Roger Federer v Andy Roddick as it happened ...
Game-by-game commentary as Roger Federer beats Andy Roddick in the men's singles final at Wimbledon.
www.telegraph.co.uk/sport/tennis
guardian.co.uk
Wimbledon 2009: Roger Federer v Andy Roddick - as it happened ...
Roger Federer claimed his record 15th grand slam in an epic. Scott Murray was watching ... Wimble-don 2009: Roger Federer v Andy Murra Roddick - as it happened ...
www.guardian.co.uk/sport/2009/
Wimbledon Men's Final: Roger Federer Beats Andy Roddick in ...
The Epic Story Of The Roger Federer - Andy Roddick Wimbledon ...
Federer-Roddick rivalry
Selected results


Refine search
New search
grand slam
epic
roddick
sports
federer
andy roddick
final
tennis match
2009
roger federer
wimbledon
horizobu
exploration engine
beta
More
Results
Select all
Selection
Delete all
huffingtonpost.com
melayubolehla.blogspot.com
tennisnewsonline.com
fuzzyyellowballs.com
telegraph.co.uk
guardian.co.uk
Roger Federer vs Andy Roddick — Wimbledon 2009 Final
Wimbledon 2009: Roger Federer v Andy Roddick as it happened ...
Game-by-game commentary as Roger Federer beats Andy Roddick in the men's singles final at Wimbledon.
Wimbledon Men's Final: Roger Federer Beats Andy
Federer-Roddick rivalry
Selected results
Delete results
delete result
save result
select all results
move result here to delete
move result to selected results

horizobu – explore, choose and find

horizobu is a new hybrid search engine based on well known algorithmic concepts and new community concepts. Horizobu is different from common search engines in different aspects:

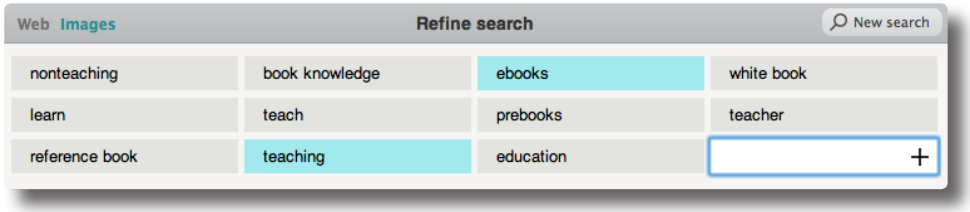
- horizobu widens your horizon and offers you new surprising insights.
- horizobu assists you in refining your search and reduces the chances of missing important information.
- horizobu interacts with you in your language and assists you in making right choices.
- horizobu suggests you search terms and you have only to click.
- horizobu does not require much keyboarding and thus making it suitable for mobile devices as well.
- horizobu offers a new search engine interface paradigm and allows you to collect and self-manage your results.
- horizobu invites you to share your search experience with your friends over the world.

horizobu is more than a search engine,
it's an exploration engine!

5/12

horizobu's user interface

horizobu's aim is to start a conversation with the user. For each search query, horizobu is generating approximately 3000 tags. horizobu is selecting about a dozen tags out of those 3000 tags using a TagRank algorithm. The selected tags are then offered to the user as possible refinements for her search query. As a result, the user doesn't have to invest a lot of time in carefully selecting her search terms but can rather start a conversation with the search engine and iteratively and interactively refine the query. The picture below shows the tag cloud for the query *ebooks teaching*.



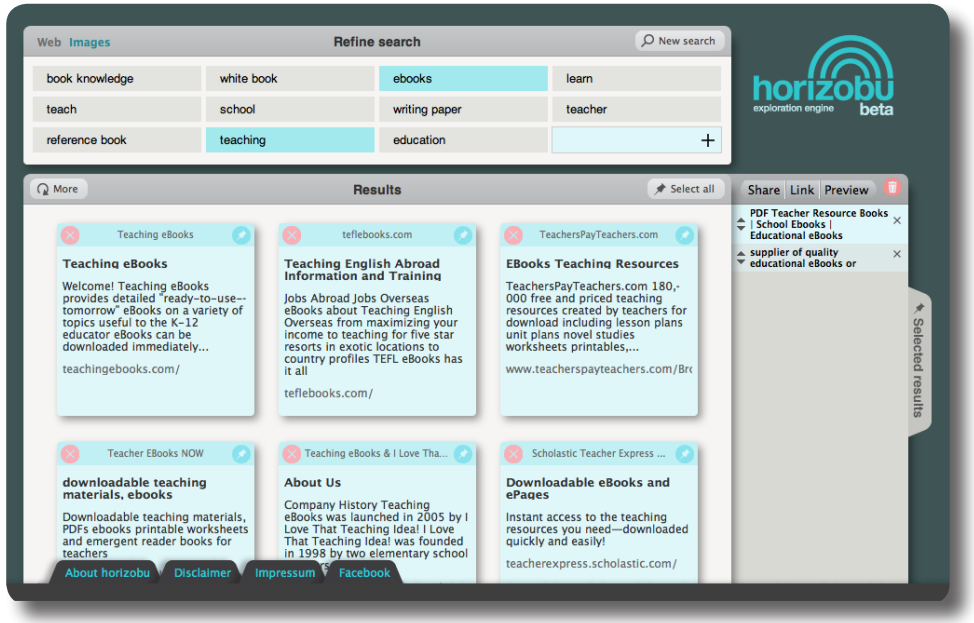
Tag cloud for the query *ebooks teaching*

By using context-based suggestions for possible search terms, horizobu is informing the user about other relevant contexts of a search query. These contexts are often not mentioned in the case of precision oriented search processes. Therefore, horizobu is not only helping with search but as well expanding the horizon of the user. This is especially relevant whenever the user doesn't know what he doesn't know :-).

Most currently used search engines show the results in a fixed ranking which cannot be modified by the user. It's neither possible to remove result nor save individual results for later use. From a usability point of view, a fixed ranking is contrary to how we search and select items in our daily lives. When looking through the mail, looking at vacation catalogues, or buying new glasses, we sort the items into two categories: good and bad. Vacation catalogues for destinations that we are not interested in are thrown into the paper pin immediately. Catalogues of destinations that might be

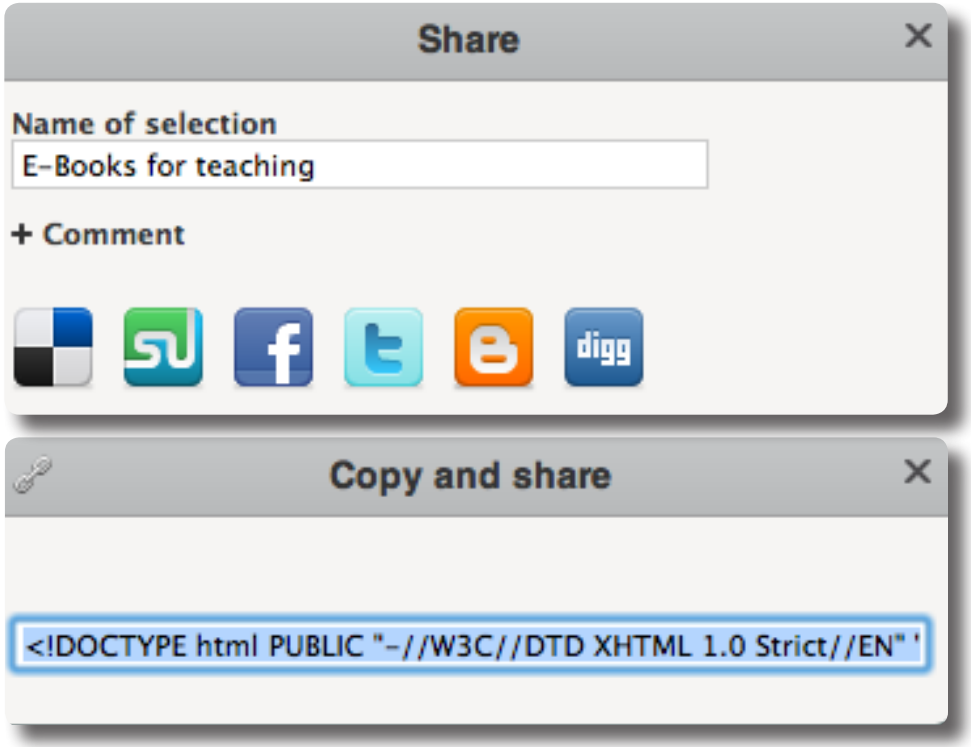
more

relevant are put aside for later consideration. We repeat this procedure iteratively until we have found the destination of our liking. horizobu's user interface is based on this paradigm: search results can be deleted or put onto a pin wall so that they can looked at again later.



Search results can be put onto a pin wall

A user can further save a pin wall for later use or for sharing with other users.



Save a pin wall for later use or for sharing with orher users

Next steps and outlook

In recent months, horizobu made stack.horizobu.com available. The same principles as for the main horizobu search are applied to the various Stack Exchange properties and should help programmers find answers. It is planned to offer access to additional vertical collections; e.g. targeted towards lawyers.

Besides the development of the system that is powering horizobu, we have as well executed an evaluation study with 75 people. We investigated whether the suggested tags are really helping the users refine their search queries. The study showed that users with good to high information retrieval skills are better supported by horizobu compared to e.g. Google. Users with lower information retrieval skill preferred the significantly simpler user interface of Google and are not bothering about information that they will miss. Rarely used was horizobu's sharing functionality. It seems that it takes time for users to understand that sharing of a search process can be beneficial to them.

It's not the goal of the people behind horizobu to be in competition with Google. Our priority is to explore new approaches in information retrieval. We are convinced that the quality of the tag suggestions and the results could be improved substantially with additional resources.

horizobu's concepts

The most commonly used methods for providing access to a collection of information are based on two principles. Firstly, the categorization of information by experts (e.g. libraries) and secondly, the use of an algorithmically generated index (e.g. traditional search engines such as Google).

Systems based on categorization use a centralized approach. Subject experts determine which keywords and categories are to be used. The categories provided by e.g. Yahoo are an example for such a system. To a certain degree, information available on wikipedia is as well categorized by subject experts. An advantage of a system based on categorization is that a subject expert has checked the content of a document and has assigned a categorization based on human brain power. But at the same time, this approach can as well be a disadvantage. Subject experts, e.g. librarians, will categorize documents based on their subject knowledge and often use subject specific terms. In order to be able to use such a categorization, the user has to have knowledge of the subject specific knowledge that the person creating the categorization, e.g. the librarians, has.

Based on that observation, horizobu is using the tagging of content by users done on Web 2.0 services; amongst others data from services such as YouTube, flickr, or delicious. Users will use their language to tag content and are therefore closer to the normal search engine user than e.g. a librarian. During the development of horizobu, we designed and executed a study which showed that user generated tags are highly relevant for information retrieval. There is high probability that a tag of a relevant document is matched by a query term. Firstly, because tags are often not in a subject language but rather in the language used by a user in her daily life. Secondly, for many social tagging services, there are many tags available for a single piece of information. One such example are books: A classic library index will contain only a couple of keywords whereas the same book will be tagged with many different keywords in a service such as LibraryThing.

horizobu basically used the query terms as tags to retrieve information from social tagging systems such as YouTube, delicious or flickr. horizobu considers at the same time result pages and checks whether those pages have been tagged

more

in social tagging systems. Amongst others, those tags are then used to provide suggested search terms to the user which he can then use to refine her search query.

horizobu is as well using the publicly available wikipedia data. horizobu generates a set of tags for every wikipedia page. Those tags are e.g. based on markup such as titles or bold text, on meta data of images or on external links. horizobu analyses as well the relationship between different wikipedia pages; e.g. two pages linking to each other are likely to be semantically related and therefore the concepts represented by the two pages are connected. Further, wikipedia disambiguation pages are playing an important role. The relationships presented by disambiguation pages can be used for the exploration part of horizobu; horizobu suggests query terms that are related to other aspects of the initial query and therefore starts a dialog with the user.

Wikipedia articles contain external links which are, in most cases, carefully selected and contain relevant high quality information. horizobu is using those links and increases the relevance of those pages and their associated information accordingly. Compared to the PageRank algorithm, horizobu puts more emphasis on the semantic tagging by the users and less on the popularity of a web page.

Algorithmic search engines such as Google are based on an full text index. For every given document, more or less all words are put into the index. Therefore, the index is seemingly complete and it looks as if such an index is better than a folksonomy. The completeness of the index can at the very same time be a disadvantage. Even though a word might occur in a document, it might not be a accurate description of the document. And there are many words, such as generic terms, that are not occurring in a document and are therefore not contained in a full text index. Tagging could be seen as a manual way of creating an index. Contrary to a full text index, the words in the index are created by humans. There will be less words in the index but most of those words will have a higher descriptive value. Additionally, generic terms and terms commonly used in a subject area will as well be part of the index created by a folksonomy. Even though a folksonomy of a social bookmarking service might be relatively small in size compared to a traditional index, it can still be better than a traditional full text index. As a consequence, horizobu is based on both a traditional algorithmic search

more

based on a full text index and a folksonomy created by the communities of social bookmarking services. horizobu is using Yahoo BOSS as the algorithmic part of the search process and web 2.0 services such as flickr, YouTube, and wikipedia for the folksonomy part.

horizobu uses new principles as well for the ranking and the presentation of the results. There are many studies showing that users are only looking at the first couple of results. Therefore, it is explicitly our goal to not provide 1000 hits but rather restrict ourselves to 50 results per query. In order to improve the ranking of the results, we consider as well behavior of users on horizobu. Such behavior can be visiting of results, saving results for sharing or deleting of irrelevant results. Further, the ranking is adapted based on popularity on e.g. delicious or relevant wikipedia articles get a higher ranking.

The history of horizobu and the people involved

The idea, on which horizobu is based on, was developed in 2008 based on an analysis of the major flaws of the existing information retrieval systems (such as Google). The following two flaws were identified:

- Search engines such as Google have access to more and more information. The results presented to the users are customized based on geographic location, data from social networks, search history, and other user-specific factors. This leads to search results that are better tailored towards the need of the individual user. Said differently, the user sees only information that reinforces his own opinion; contradicting information is filtered out and therefore the user is presented only a partial picture.
- For a long time, the information retrieval research community has identified users as the bottle neck. Often, users are using an insufficient amount of query terms or query terms that are not specific enough. As a consequence, search engines have often not enough information to define a result set that is relevant for the user. Even though there has been a lot of progress in the area of search engines, search engines are not getting into a conversation with their users.

A German beta version went online in fall 2010 and we have been continuously improving it since.

www.horizobu.com