

An Interview with Danica Brinton of Ask.com

By Sébastien Billard (s.billard@free.fr)

Originally published on <http://s.billard.free.fr/referencement>

SB : Hi, first, thanks for accepting answering some questions, could you introduce yourself to readers ?

DB : You are absolutely welcome. It's a real pleasure, Sebastian. My name is Danica Brinton and I head International Product Management and Localization for Ask.com.

SB : What distinguish your blog search tool from others ?

DB : We built a system that delivers superior results and high quality content with low spam content and high level of relevance. And we did it in an extremely intuitive way.

We feel that crawlers used by standard search engines fail to expose the full blogosphere. Syndicated content presents search engines with a unique challenge: capturing the full diversity and freshness of the blogosphere, while ensuring top-quality relevance. Search engines that merely extend Web search techniques to syndicated content, by simply crawling blogs and other sites, are doomed to fail this challenge. Unlike the static Web, the blogosphere evolves too quickly for robust link structures, the life-blood of crawlers, to develop sufficiently for use in discovery of new content. As a result, crawlers, and the search engines that use them on the blogosphere, invariably miss important information, or look to other methods (such as pings) that are overly susceptible to spam.

So, instead of crawling, Ask Blog & Feed Search harnesses the subscription data of hundreds of thousands of real people who use Bloglines, the #1 online feed reader, to create our search index. In the absence of a mature link structure, people provide the best way to discover the freshest, highest quality feeds -- information that isn't exposed to crawlers. In addition, this "collective human intelligence" provides a natural defense against spam, as people typically do not subscribe to low quality content.

Because Bloglines is the largest and longest established major blog reading community online, Ask Blog & Feed Search also has the most robust index of content on the Web: articles are indexed from 2001 through five minutes ago (or less). New posts are added at a rate of four to six million per day, with a total index in excess of 1.5 billion articles, with 4 to 6 million added every day.

On top of this superior index, Ask Blog & Feed Search applies our unique, world-class algorithmic search technology, enhanced by data from the Bloglines community, to deliver unrivaled relevance.

We believe that our product offers very instinctive and quite necessary tools. Ask Blog & Feed Search lets you search or toggle through three types of results :

- Posts : Relevant posts (or articles) that match the query topic. Over 1.5 billion posts have been indexed.
- Feeds : Relevant feeds that match the query topic. (Denoted by a feed's favicon where

available.) Over 2.5 million individual feeds, with subscribers on Bloglines, have been indexed.

- News : Relevant posts specifically from a sub-index of approximately 7,000 news sites.

Sorting works by Most Recent, Popularity and Relevance. Within each search type, you can sort in one of three ways to find useful information :

- Relevance : Based on a combination of Date and Popularity. This is the default option.
- Most Recent : Sort by date.
- Popularity : Popularity is determined by a combination of subscription, link/citation, and ExpertRank community data.

Preview feeds by simply mousing over the Binoculars icon within your search results. Binoculars is a patent-pending preview technology that enables you to quickly preview feeds before clicking-through.

After finding relevant results, Ask.com makes it simple to manage information directly from the Ask Blog & Feed Search results page : Use the Subscribe drop-down to subscribe to feeds not only in Bloglines but also other services, including Google Reader, NewsGator or Netvibes. Use the Post To drop-down to clip the search result directly to services like Bloglines, Blogmarks.net, Linkedfeed or Mesfavs.

You can set up a persistent search based on the current search topic and find out almost instantly when new content appears on the blogosphere matching your topic. You can take this subscription with you, as well, by selecting your favorite Web service, including Bloglines, Google Reader, and MyYahoo.

Our Blog Related Search provides related feeds when searching for posts. Appears down the right side of the search results page to help guide you to additional relevant content.

You can save your blog search results to MyAsk.

Our Advanced Search allows you to hone queries with a variety of options, including the ability to select one or more of the 20 supported languages. On Ask, the Advanced Search feature is exposed through seamless page integration that drops, in sliding fashion, vertically into place.

(I hope you don't mind my long-winded answer) :)

SB : Can you explain briefly the ExpertRank algorithm, and how it is used concerning feed search ?

DB : ExpertRank is a unique ranking algorithm that relies on communities and clusters in search. To rank an item, it is not enough to know the link structure. Link structure can be artificially manufactured. We rely on authoritative information about those links.

SB : How does the blog search collect feeds ? Is it by crawling the web ? By using the subscriptions of Bloglines users ? A mix of both ? How bloggers and content producers can make sure their feeds are indexed ?

DB : Bloggers and content producers need to simply subscribe to Bloglines in order to add their content to our blog search index. Quite simple. :)

SB : How does the blog search engine determine the best flux displayed on the right of the screen ? Is it based on the number of subscribers in Bloglines ?

DB : We observe the number of subscriptions but more importantly the links, citations and their value. Then, we add our special sauce. :)

SB : Concerning the search of feeds, how does the search engine determine if a flux is relevant for a keyword ? Is it only based on title, description, and content of the feed at a given time ? Or is there some analysis ran to determine the general themes of a feed ?

DB : I believe I answered your questions above already. In quick summary: user votes, citations, links, content and Expertrank.

SB : The Blog Search doesn't return the same results for a word with and without accents (see "referencement" and "référencement"). Is it a bug or feature ? Don't you think it should return the same results as the omission of accents is 99.99% of times laziness or misspelling ?

DB : I appreciate your feedback. I will look into this right away. In general, we are very careful about normalization and often find that a user intent may be different with varying accent use but you are right: a lot of the time it is a result of an English keyboard or speed of typing. If you have any other feedback on the product, please, do not hesitate to let me know. Our French site is in Beta right now and feedback from an expert like you is invaluable.

SB : I noticed many Digg-likes websites indexed in search results (Tapemoui.com, Fuzz.fr...) But this kind of services only list links to resources, they are not resources themselves. Many times, they use a blog post title, letting the user think the resource is behind the Ask's link, whereas it is one more click away. Do you consider this a problem for relevancy, and have you some solutions for it ?

DB : I agree with you again. We are carefully sorting the content that you pointed out. There are some challenges there and I'm sure you can guess where they are coming from.

SB : Have you some algorithms or do you use human intervention to avoid the indexation of RSS feeds that are only tools, not information ? I am thinking to RSS feeds from Wikipedia that list changes to pages, for example.

DB : We are algorithmically controlling this content. I hope you don't mind if I keep the rest a secret. :)

SB : I noticed some links in SERPs that looked like tracked (beginning with wzeu.ask.com). A parameter of the url is named "ip". Is it some quality evaluation ? Or is user tracked for personalization of results ?

DB : Great spot! But I'll keep the answer confidential. (hope you don't mind) :)

SB : A last question, concerning the web search engine : When the Zoom feature will be available for France ?

We add our core features to our international sites constantly. Zoom is one of the features that we'll launch on our international sites once we are out of Beta.