

7 Technical Sources

Although most of the following sources could be considered precedents, I decided to make a distinction according to visual depiction. Even though the listed websites represent strong conceptual precedents to my thesis, its outlined structure denotes a stronger and more tangible technical source for its future development. Also, in opposition to visual analysis/representation, most of these resources display only text-based information, regarding *blogs* connectivity and topics popularity.

7.1 Blog Engines

BlogPulse, **Bloggz**, **Technorati**, **Popdex** and **Bloglines**, among others, are *Blog Search Engines*, similar to **Google** or **Yahoo**, but restricted to the *blog* community. Besides offering a free *blogosphere* search, they list the most popular *blogs* (by daily number of visitors or inbound links), the most common search inputs, and the trendiest topics (by recent added links/quotes on daily *blog* entries).

Some of these websites offer additional information, such as: the evolution of a key sentence/word in static diagrams, shared links between *blogs*, news/words popularity, key people, discussion threads, neighborhoods, and *blogrolling*. Most of these features are text based, listed according to date, popularity, or a specific ranking. Many of these resources also provide RSS feeds of their content in XML format, which can be read by browsers or *newsreaders*, applications similar to email programs that interpret the feeds and list them by title, with a small descriptive paragraph and a link to the source.

Regardless of the interesting features of these services, on capturing the trendiest bustle in the *blogosphere*, most of them present solely textual rankings, which are difficult to compare with additional factors and derive further conclusions from an eventual information overlapping. Besides the lack of visualization, that would help better understand this complex network of dependencies, most of these services are only worried in capturing the momentary fad. As one can read in **blogdex.net** homepage, regarding the constantly updated list of websites displayed there, they represent “the most contagious information currently spreading in the weblog community”. Most of these services have disregarded the understanding of the evolutionary process of information contagion and the dynamics of these diffusion patterns.

Here are some of the most prominent *blog* services:

- ∴ <http://www.technorati.com>
- ∴ <http://www.blogpulse.com>
- ∴ <http://www.bloogz.com>
- ∴ <http://www.blogdex.net>
- ∴ <http://www.popdex.com>
- ∴ <http://www.bloglines.com>
- ∴ <http://www.daypop.com>
- ∴ <http://www.blogstreet.com>
- ∴ <http://www.blogwise.com>
- ∴ <http://www.feedster.com>

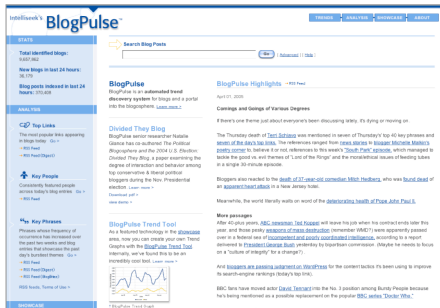
7.2 Blogviz Data

All the data used in Blogviz was obtained from three main blog services. The daily topics were collected through **Blogpulse.com**. The correspondent blogs for each topic were taken from **Blogdex.net**. And finally, the number of inbound links for each topic was obtained from **Technorati.com**.

The main reason why this procedure was spread over 3 different sources was that none of them individually congregated all the needed elements to use in Blogviz. In the beginning I was expecting to work with a single data source, but further along in the process I concluded that each source had a specific core asset that was particularly useful for the development of the project. I believe this process was also positive since it took a more pluralistic approach not constrained by the limitations of a single source.

The data was collected through the months of April and March 2005. I started by building a personal database, with the use of ColdFusion, MySQL and Microsoft Access. It took some time to have the database up and running, but after the structure was created and operational, all the effort was concentrated on inputting the data manually in the new datasource.

7.2.1 Blogpulse (<http://www.blogpulse.com>)



“BlogPulse is an automated trend discovery system for blogs. (...) BlogPulse applies machine-learning and natural-language processing techniques to discover trends in the highly dynamic world of blogs.”

BlogPulse.com is basically a portal into the world of blogs.

Here are some of its features:

- A search engine for blogs
- A daily list for blog content (top links)
- A look at real-world trends as reflected through blogs (static diagrams)
- A showcase, seen as virtual sandbox where researchers bring ideas, tools and gadgets for blogging

The key advantage of Blogpulse is that it saves its lists of daily top links in a large archive, easily accessible by the general public. Each html file can be searched by date of occurrence. So imagining that in November 24, 2005, I would be interested in knowing what were the most popular links among *weblogs* in February 08, 2004, Blogpulse would allow an immediate access to this information. This feature is of extreme relevance to the development of Blogviz. Its use not only allowed this thesis’s contention to expand, but also, will facilitate its continuous development.

In spite of the utility of this feature, Blogpulse is slightly limited when it comes to the description of adopting blogs (or citations) for each topic. The list does not contain a time of “adoption” for each blog, and even the overall perpetuating days are not easy to perceive. This is why blogdex.net comes as a second data source for Blogviz.

7.2.2 Blogdex (<http://www.blogdex.net>)



Blogdex is a research project developed at MIT Media Lab intended to track the diffusion of information through the *weblog* community. It lists a sequence of links considered “the most contagious information currently spreading in the *weblog* community”. Cameron Marlow, a PhD candidate at MIT Media Lab, is the key protagonist behind this project.

“Blogdex crawls all of the weblogs in its database every time they are updated and collects the links that have been made since the last time it was updated. The system then looks across all weblogs and generates a list of fastest spreading ideas. This is the list shown on the front page. For each of these links, further detail is provided as to where the link was found, and at what time.”

The core usefulness of Blogdex in the development of Blogviz is that it has one of the most organized sorting systems of adopting weblogs for every topic. Whatever topic a user searches, besides the front page list, Blogdex displays a list of adopting blogs, efficiently organized by date and specific time of adoption. This allows a significant understanding of the topic evolution by tracking the exact time of “contagion” for every blog.

7.2.3 Technorati (<http://www.technorati.com>)



“Technorati is a real-time search engine for the blogosphere. Technorati tracks the number of links, and the perceived relevance of blogs, as well as the real-time nature of blogging. Because Technorati automatically receives notification from weblogs as soon as they are updated, it can track the thousands of updates per hour that occur in the blogosphere, and monitor the communities (who's linking to whom) underlying these conversations.”

Because Technorati is probably the current largest blogosphere search engine, it was used to measure the broad scale of popularity for each weblog collected in Blogviz’s database. The popularity of a blog represents the number of inbound links it has, or in other words, the number of blogs that link to it.