Pfizer

Blog Analysis of European Healthcare & European Pharmaceuticals

SEHTA Business Intelligence Team

Emanuela Todeva, Donka Keskinova, David Parry

Leatherhead, September, 2007[*]

---

[*] This report is intended as a basis for discussion. While every effort has been made to ensure the accuracy of the material in this report, the authors and the sponsors will not be liable for any loss or damage incurred through the use of this report.

Introduction

This report provides the findings from a research project commissioned by the Business Innovation Unit in Pfizer Ltd. The SEHTA team was asked to assess the emerging body of information being published in web logs (blogs); one of the emerging channels at the heart of the rise in social media now commonly referred to as Web 2.0 (Enterprise 2.0 in a business context). The hypothesis behind this request was that the rise of blogs as a rich information source may provide a challenge to the traditionally held view of a 'key opinion leader' – could an anonymous blogger have as much influence over public debate as a recognised scientific expert? *The project's goal was to carry out initial research into the area of health blogs, and provide recommendations as to future areas of focus for research.*

Analysis of the Blog-space: Key Observations

- The blog-space is a dynamic configuration of the internet with continuously changing entries and exits.

- There is very little interaction and public response compared to the volume of text that is broadcasted in publicly accessible blogs.

- There are many substantially different formats that are in use, and it seems that there is no dominant pattern of format emerging. Most blogs have options for enabling comments and other interactions, such as tagging or emailing an article. However, their classification as blogs and/ or selection by blog-search engines is often due to technical features such as meta-tags in the HMTL code of a web page.

- The majority of blogs in the area of European Pharmaceuticals and European Healthcare are technical and organisational experimentations and explorations that aim to broadcast information. There are two main types of blog-news - *generalist* news (blogs established by the main media with publications or specialised sections on healthcare and pharmaceuticals), and *specialised* news (blogs established as specialised sources of information on medicine and healthcare). There are some 'community-type' blogs that stir community interactions, and some personal blogs – as individual attempts for expression of opinion.

- Almost all blogs have included in their registration entry a copyright claim, which prevents from a free use of the text for analytical purposes. For the purposes and duration of this project – no permissions from copyright holders were obtained, and hence, we were not able to test our methodology for direct text analysis.

- Multiple blog search-engines are available. They vary substantially in the results and ranking of the results that they retrieve from the world-wide-web.

- The majority of blog postings have not received any comments from the audience. If comments are posted, they often evolve in one thread that follows up upon one article and they are written within one to two days from the original post.

- Many blogs use automated facilities for organizing and structuring the information, e.g. via time-based archiving of posts and tag-based aggregation.

- The main volume of blogs has emerged from the beginning of 2007, and this project is very timely – to map the beginning of a new process of on-line interactions and the development of an interactive public space for information-retrieval.

Executive Summary

- The blog analysis revealed that the interactive public space for information retrieval via blogs and web2 technology is rapidly growing. Most of the technologies that support the registration and search of blogs are still at their developmental stages, and major changes or domination of particular technologies are quite possible.

- We have developed an original methodology for searching the blog-space, for building databases with registered URL-pages, for in-depth analysis of the content of URL-pages, blog-content, blog-structure, as well as mapping the relationships between key actors in this public space. We have employed all these methodologies for the completion of the current report with exception of our blog-text analysis – due to copyright issues. The details of our methodology are described in the methodology section further in this report.

- For the purpose of our comprehensive analysis we used two frameworks: one included the entire semantic field of European pharmaceuticals and European healthcare; and the other – the leading 19 pharmaceutical firms from a list selected by the client (see Appendix 2).

- All pharmaceutical companies included in our search have a presence in the blog-space – with exception of Lacer. Large firms attract a lot more attention, and the reference to Pfizer is dominant (279 URL-pages for the period up to July 2007), followed by GlaxoSmithKline (205 URL-pages), Novartis (194), Bayer (159), Sanofi Aventis (156), Eli Lilly (146) and the rest.

- The mapping of the entire blog-space for European pharmaceuticals and European healthcare has an interconnected core of 36% of all URL-pages and a large periphery of URL-pages related to a single pharmaceutical company. 31% of the URL-pages (or 86 URL-pages) refer only to Pfizer and no other pharmaceutical firm from our selection. Although Pfizer appear to occupy a space fairly at a distance from other pharmaceutical firms, it is also connected to Johnson & Johnson, Merck & Co, and Takeda, particularly on issues related to 'healthcare', 'healthcare system', and 'health policy'. In addition, Pfizer appear to be strongly connected to blogs such as: Pharmalot, Pharmasia News, RxBlog, Talk: Med, Canada's shame, Computer Monkey, Forward in reverse, and Google-Sina Medical Health – among others.

- Although Pfizer is a dominant actor in terms of volume of blogs in which it appears in reference, however, it does not appear to have a distinctive profile. It appears rather generalist – in the semantic fields of DRUGS, HEALTH, and DISEASE. This is in contrast to some other pharmaceutical companies that appear closely associated with a particular treatment area and health issues.

- The relationships between URL-pages in the blog-space are still very rare and emerging. One of the blogs that has created fairly dense internal and external links is Garbage-garbage, and this area requires further attention.

- The analysis of the semantic fields of HEALTH and DISEASE reveal emerging threads of inter-related issues, as well as semantic distances, such as: close proximity between medicine, patients and healthcare trust – on one hand, and private hospitals, medical tourism, and healthcare system – on another. Disease areas such as metabolic, gastrointestinal and respiratory appear quite interconnected and central to the public discussion.

- Large pharmaceutical companies appear to have a broader impact on the blog-space dominating articles and publications, while small firms appear most often in the shadow of another large

pharmaceutical firm. In this context Pfizer's associations with Teva Pharmaceutical, or Wyeth is visible.

Methodology and Selection Criteria

For the purpose of this consultancy project we used a comprehensive methodology for blog analysis developed by Todeva and Keskinova. The methodology comprises of the following main steps:

1: Development of the selection criteria – Downloading text from the news on European healthcare and European pharmaceuticals – broadcasted between January-June 2007; text analysis (with proprietary software) of the news and identification of key words as 'search-key-words'; grouping of the key-words in 6 distinctive groups (health, drugs, diseases, industry, regulation, region). The key-words for the selection criteria are listed in Appendix 1.

2: Selecting a blog-search engine – Initially a number of blog-search engines were used in order to evaluate which one can provide the better capabilities for blog analysis. Two search engines were explored with more depth – Technocrati (http://www.technorati.com) and Google blog (http://blogsearch.google.com).

3: Search string – We formulated search queries that combined positive and negative filters with Boolean operators such as AND and OR. The positive filter contained three components:
- the scope of the research (pharmaceutical / healthcare)
- geophysical relevance (Europe, UK / England, France, Germany, Spain)
- one of the selected key words (from our selection criteria – see Appendix 1).

The negative filter contained *Viagra*. Furthermore, we constrained the search to English pages as well as the blogs created or modified after January 2006.

4: Provisional typology –In order to filter blogs by content and impact and to sample them we developed a provisional typology which was used at the initial stage of building the database. This included: community discussion, news discussion, institutional discussion, 'private show'. This typology can be used further in evaluation of the context of on-line interactions, and a full-scale text analysis.

5: Building of the databases – For the purpose of this analysis we explored different methods of sampling the blogo-sphere. Our first database (DB1 – conceptual) was generated with a selection of URL-pages using all key words (155 cases). Preliminary analysis of the results showed that with this sampling there is too much noise and results are not very meaningful. The second database (DB2 – formal) was generated as a representative population for the blogs that contain at least one key-word from our search criteria, where we constrained the sample to the key-words with a name of a pharmaceutical company (Pfizer, Glaxosmithkline, Sanofi Aventis, Novartis, Hoffmann La Roche, Astrazeneca, Johnson & Johnson, Merck & Co, Wyeth, Eli Lilly, Bayer, Lacer, Bristol Myers Squibb, Shire Pharmaceuticals, Chiron Corporation, Chugai, Takeda, Teva Pharmaceuticals, Ranbaxy). Using these precise key-words, we were able to assemble all blog-entries, and to build a comprehensive database of the full population of blogs that comply with our selection criteria[i]. To enrich the scope of the database we included two phrases as key-words – 'European pharmaceuticals', and 'European healthcare'.

6: Collecting data from the internet – We downloaded all blog-pages that appeared accessible using Google Blog search engine. On each of the selection key-words we made a note for the total number of results in order to track the population. We noted that the blogo-shpere is a highly dynamic space. This means that individual blogs can be shown as a duplicate result by the search engine, as well as suddenly

---

[i] This statement has to be interpreted in the context of the dynamic nature of the blog-space, where blog-accessibility varies continuously.

removed from the selection. For our DB2 we copied all entries that were identified by the search engine after filtering the majority of duplications by the search-engine itself. The automated reduction was 75%, where the population of blogs was reduced from 11824 to 2995 by the search engine itself (see table 1a below and table 1b with full details in Appendix 2).

*7: **Structure of the raw data*** – The main data automatically collected per blog included the following: *Page Title* (where the selection criteria are fulfilled); *URL-Page*; *Month* (of registration); *Year* (of registration); *Blog name*; *Blog URL*.

*8: **Cleaning of databases*** – After collecting the details per blog, we cleaned the database – to isolate the data from the noise and from duplications. For this purpose we used observation techniques, and formal techniques based on proprietary software for URL searches. The cleaning of the database passed through the following stages: cleaning of duplicate URL-pages; cleaning of 'empty-URL-pages' with size < 2 KB information; cleaning of 'shell URL-pages' that contain dictionaries, job-announcements, lists of URLs without text; and URL classifications, or adverts. The details of how this cleaning affected the database are presented in Table 1a and Table 1b in Appendix 2. According to this procedure we built a database with the full population of blogs that corresponded to our selection criteria, containing 990 entries. Out of this population we identified 357 blogs (or 36%) as a core that contain more then one of our key words, and 633 blogs as periphery – i.e. related to only one key word (please, see Appendix 2).

Table 1a. Population size

Key words from selection criteria	Total Google-search Results	The most relevant results	- less duplicates	- less 'shell-pages' = total in final database	- of which unique pages that refer only to the company / or key word
Total number of URL-pages	11824	2995	2778	990	633

*9: **Internet count of key blog-indicators*** – After initial cleaning using observation techniques, we subjected the database to multiple tests using aproprietary software for searching URLs. We calculated four additional indicators per URL, which included: *size of URL in KB*; *Cross-reference between URLs in DB* (identity of internal links); *Cross-reference to other blogs* (number of external links); and *Number of occurrences of individual key-words per page* (including double counting for occurrence in URL). Some of these indicators were used for additional filtering of the data, and the final numbers were recorded after the cleaning process was completed.

*10: **Description of the population of blogs in database*** – The population of blogs in the database was described both as a derivative from the key-words of pharmaceutical companies (Appendix 2), and as a cross-reference between URLs based on co-reference of firms (Appendix 3).

*11: **Data analysis and mapping*** – for the blog analysis we used network analysis techniques based on one-mode and two-mode graphs, where the nodes differ. For the network analysis we interpreted the 'relationships' or links between:
-   pharmaceutical firms and URLs in database (Net 1.1, Net 1.5);
-   pharmaceutical firms and key-words grouped in three groups – Health, Drugs, Disease – (Net 2.1b, Net 2.2b, Net 2.3b);
-   relationships between URLs based on cross-reference between the pages (i.e. one page referring to another by posting a URL link to it) (Net 3);
-   interconnected pharmaceutical companies (where each link between two firms represents a co-presence of these firms in a significant number of individual URL pages) (Net 4.1)
-   interconnected key words in specific semantic fields (i.e. semantic blocks) (Net.2a, Net 4.4b)
-   interconnected URLs and key-words (Net 5).
Full list of maps is provided in the presentation file 'Pfizer.ppt'.

12: ***Blog-text analysis*** – In addition to the network mapping of URLs, we designed a new methodology for text analysis of blogs. This methodology enables us to conduct an in-depth analysis of the text in a particular blog, including: tracking the life-cycle of ideas; identifying how new ideas emerge, how they are carried on and how they vanish across time and articles. Such a text analysis enables us to answer the questions - What ideas connect what blogs or comments? How comments relate to each other and/ or to the original post? Such analysis will require us to build a new database with full text where a permission from copyright holders is obtained in advance. In this database, the text has to be structured according to depth-level (main introductory page; additional URL page, comments to main page, comments to additional page, comments to comments, etc.). Additional indicators for blog-text analysis can be the number of distinct commentators, and the number of distinct comments at each level of a particular blog. Blog-text analysis is suitable for an in-depth analysis of blogs that are identified as dominant or central in a particular field.

Overview of Results

The results from the blog analysis are grouped in 5 main sections: A: mapping the blog-space of European healthcare and pharmaceuticals; B: mapping the key actors in this space; C: mapping of relationships between blogs; D: mapping of the topics on which blog-participants publish; E: mapping of the impact;

A: Mapping of the blog-space

## Net 1.1[ii] All ties between companies and page/URL
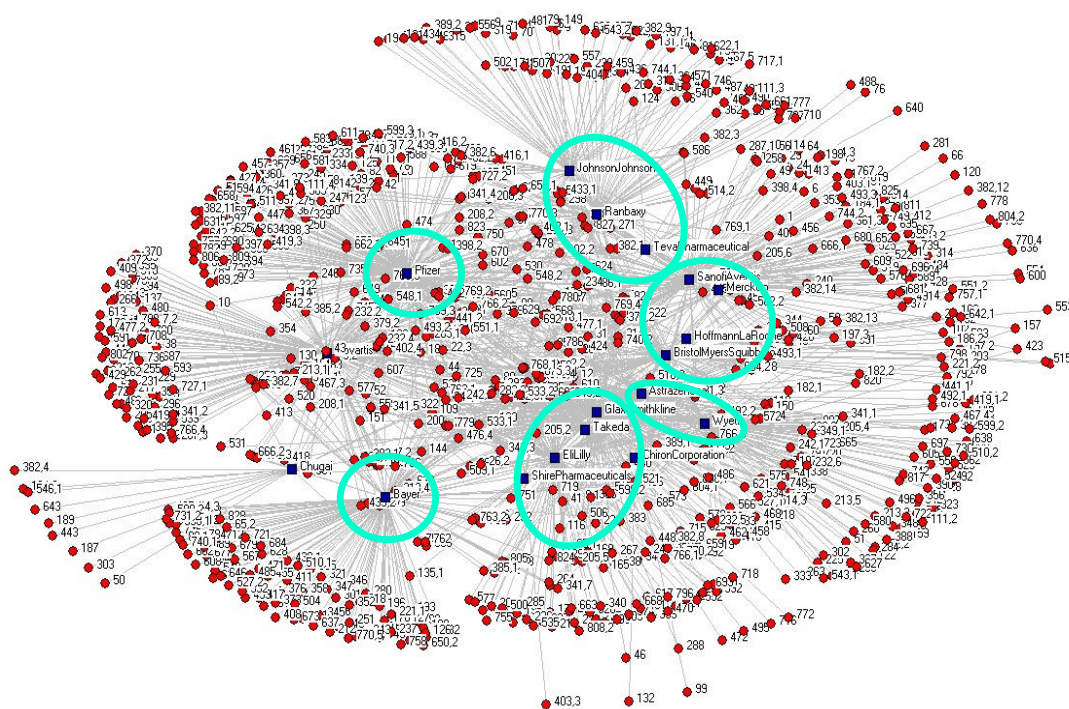
The first two maps show distribution of URL-pages and their association with a particular pharmaceutical company. Net 1.1 shows that Pfizer is quite central to the selected field, and both Pfizer and Bayer exhibit unique profiles, as there are no other companies located in a close proximity. In addition, there are three sub-groups of companies that share a significant number of URL-pages. These are: Johnson & Johnson with Ranbaxy and Teva, GlaxoSmithKline with Takeda, Eli Lily, and Shire; AstraZeneca with Wyeth, and Sanofi Aventis, Hoffmann La Roche, Merck & Co, and Bristol Myers Squibb. The strongest connections between firms and URL-pages are exhibited in Net 1.5, where we

---

[ii] The numbers of all maps correspond with the numbers in the presentation file, associated with this report.
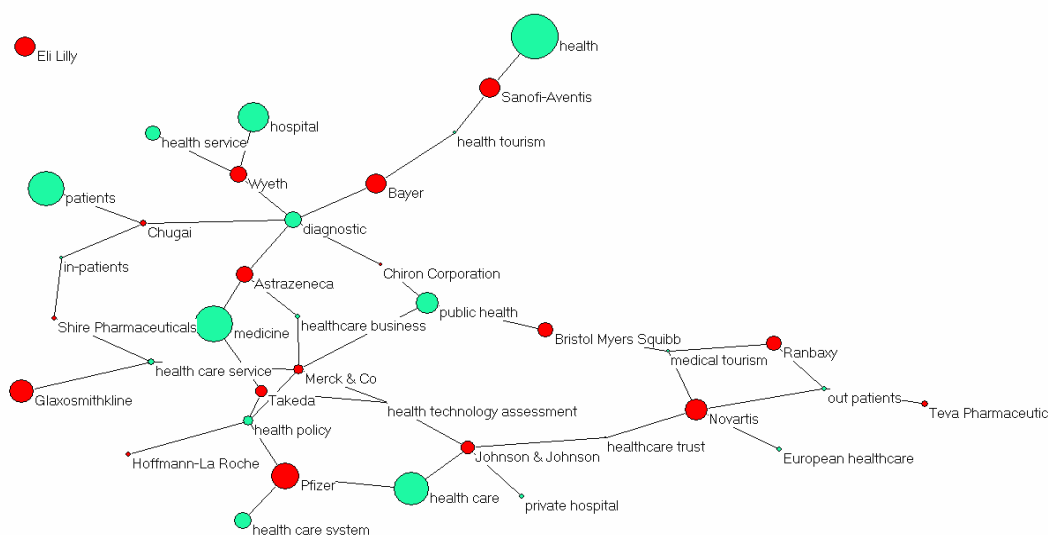
observe four blogs as the most central to this field – Pharmalot, Impactivity blog, Canada's shame, and Rx blog.

Net 1.5 More then 5 ties between companies and URL-pages (del pendants)
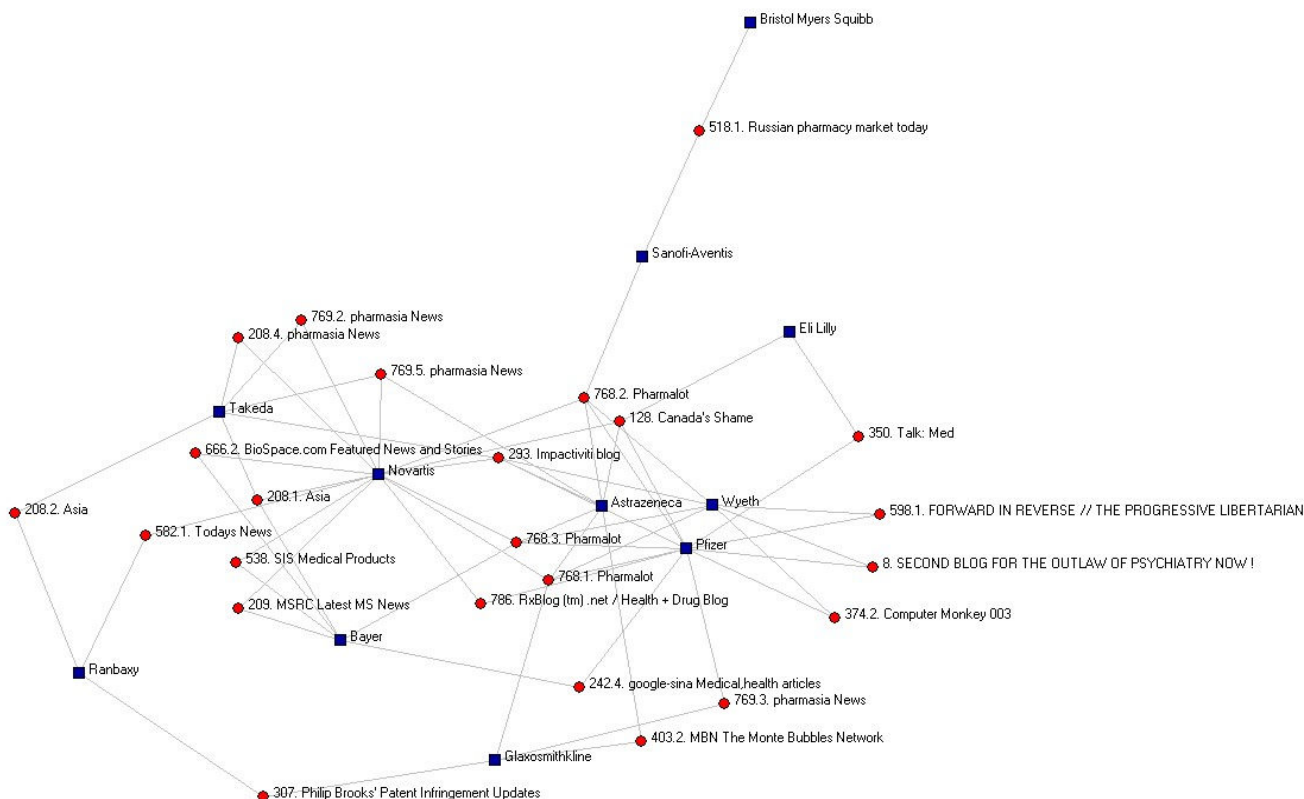
B: Mapping of the key actors

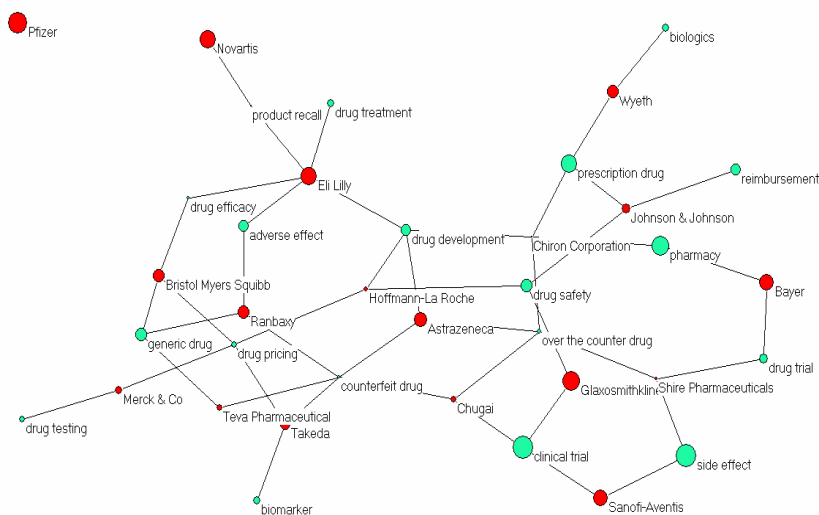**Net 2.1b Companies vs. key words in block A. HEALTH (normalised value)**

The key actors in our analysis were the pharmaceutical companies selected for the research. The mapping of key actors included analysis of the relationships between these companies and the key-words that represent the semantic field of European pharmaceuticals and European healthcare − grouped as Health, Drugs, Disease, and Regulation (see Appendix 1).

The three maps in this section identify how individual pharmaceutical companies are located in the semantic field that corresponds with European pharmaceuticals and healthcare. We have used normalised value to counteract the size effect where all firms are equal irrespective of the volume of blogs where they are referred to. In the first Net 2.1b Astrazeneca, Merck & Co, and Takeda appear most central to the debates surrounding healthcare issues, including medicine, diagnostics and public health issues. Pfizer is mostly associated with generic categories such as healthcare, healthcare system, and health policy. Eli Lily appear as an isolate in this map, which means that it has no preferential associations with any particular issues related to healthcare, but exhibits equal presence in discussions of all health issues. Pfizer has a similar position on Net 2.2b, related to discussions on Drugs, which can be interpreted as 'broad and/or indiscriminate impact'. The central issues that dominate the discussions on Drug issues are drug safety, drug development, prescription drugs and over-the-counter drugs. The discussion of side-effects is associated mainly with Shire and Sanofi Aventis. The issues of generic drugs are associated with Bristol Myers, Ranbaxy, and Teva. Wyeth leads the discussions on biologics, and Takeda leads the discussions on biomarkers.

Net 2.2b Companies vs. key words in block B. DRUGS (normalised value)

Net 2.3b Companies vs. key words in block D. DISEASE (normalised value)
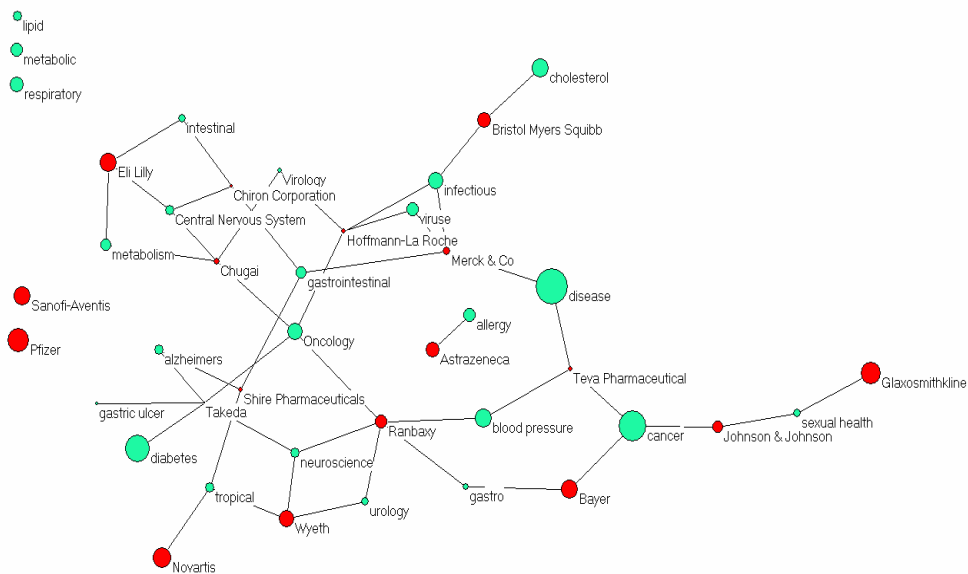

Net 2.3b shows that discussions on oncology and gastrointestinal diseases are most central across the disease space. Most of the pharmaceutical companies have a profile that associates them with particular disease – Johnson & Johnson with cancer and sexual health; Wyeth – with tropical and neurology;

AstraZeneca – with allergy; and Ranbaxy – with oncology, urology, gastro, blood pressure and neuroscience. Pfizer and Sanofi-Aventis appear as isolates suggesting that they do not exhibit preferential association to any particular disease.

C: Mapping of relationships between blogs

All blogs and URL-pages in our database are interconnected as they all represent one common semantic field – drawn by the use of key-words from the selection criteria. However, some of them are 'more' connected than others as they show awareness of each other and send links to each other. Some pages post a URL link from themselves to another blog that is considered as a relevant source of information. From the map on Net 3 we can conclude that the blogo-sphere in our field is very fragmented. There are only occasional links (cross-reference) between URL-pages forming a large number of dyads. There is only one small group of URL-pages in the centre focused on a blog 'Garbage Garbage', where interconnectedness emerges. This blog is recommended for further in-depth analysis.

Net 3. Ties between URL-pages based of internal links – node-size is equivalent to the size of the blog (KB)
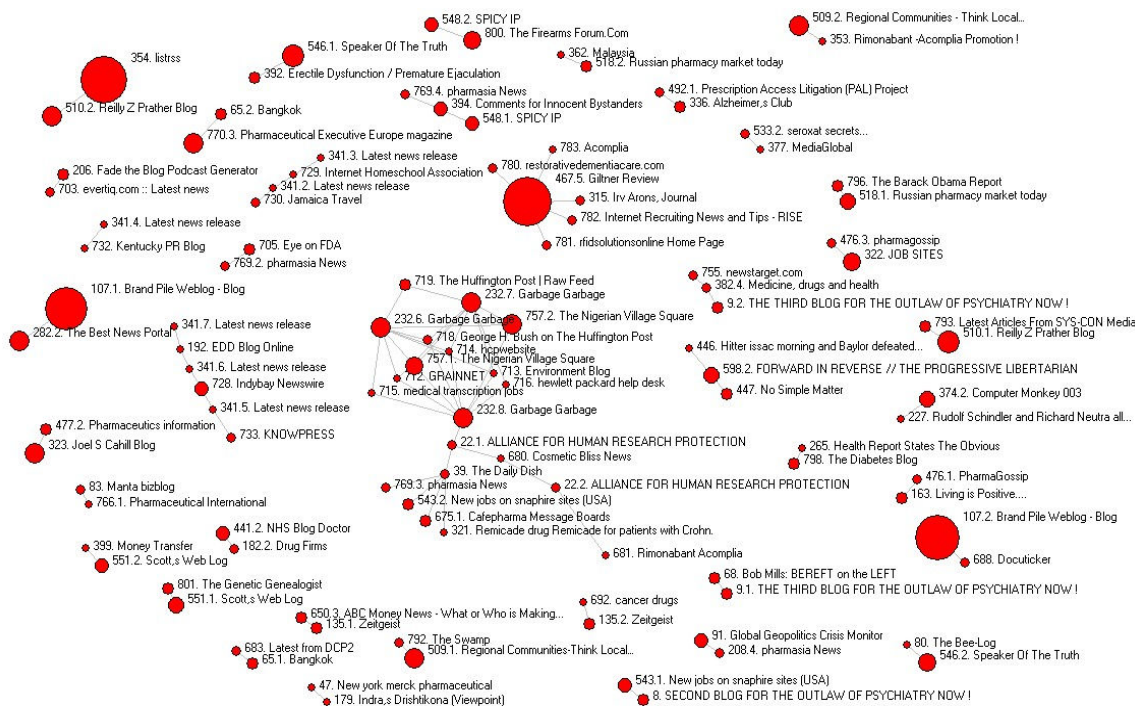
D: Mapping of the topics on which blog-participants publish

We have focused our analysis on two main areas of the semantic field – issues related to health and issues related to disease. Net 4.2a describes the leading threads in this specific semantic field. Dominant concepts are health, healthcare and medicine. The latter is further associated in discussions with patients, healthcare trusts, hospitals, diagnostics, European healthcare, and healthcare business. This thread is quite dispersed which suggests loose connections. There is a densely interconnected graph that has emerged around the core issues of 'public health', 'health services', 'healthcare system', and 'health policy'. This core has a periphery, among which are 'in-patients', 'out-patients' and 'private hospitals', which connect to 'diagnostics'. The interpretation of the dense component suggests that these issues form a homogeneous semantic field of inter-related issues covered in the URL-pages' content.

Net 4.2a Interconnected key words in the semantic field of block A: HEALTH
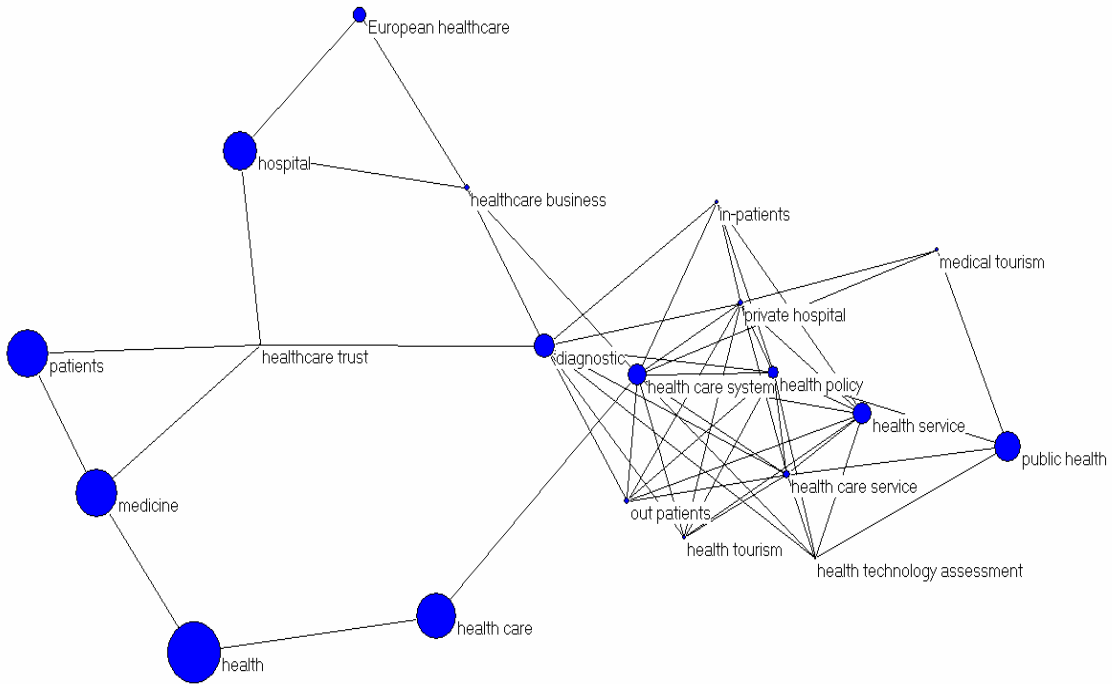
Net 4.4b Interconnected key words in the semantic field of block D: DISEASE
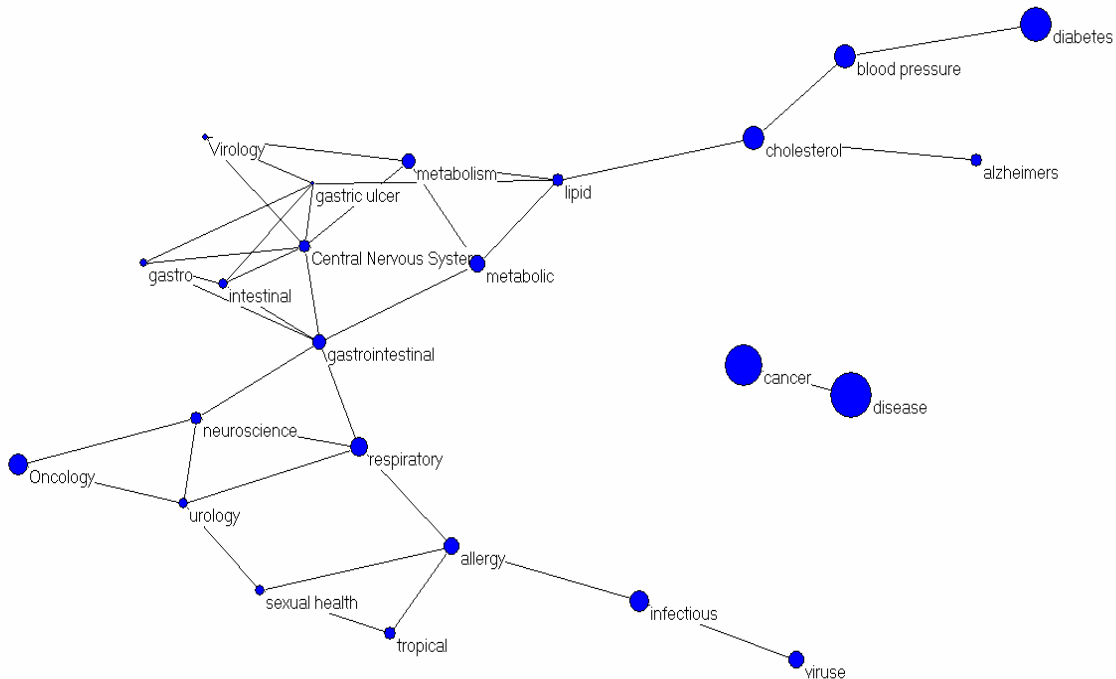
Net 4.4b Shows how different diseases connect semantically. Some diseases are quite central and form a chain (lipid, metabolic, gastrointestinal, respiratory, and allergy). Cancer and disease form a strong dyad, and show high volume of references in blogs, but in very focused discussions linking mainly the two of them.

E: Mapping of the impact

There are different ways for evaluating the impact. One of the established methods is evaluation of the centrality of blogs and URL-pages (i.e. how central and interconnected is each URL-page from a blog (Net 3.). However, this analysis reveals very limited impact – as blogs are disconnected – informing only their specific audiences, where the audiences do not seem aware of other blog-audiences, i.e. make no reference to other blogs. The number of dyadic links shows that there are only occasional links between blocks and URL-pages, but each blog entry exists mainly by itself.

We developed an alternative method to assess the impact by mapping how firms are interconnected in the selected blog space. Net 4.1 portrays the strongest links between pharmaceutical companies showing what is the individual position / impact of a particular firm in relation to all other firms.

Net 4.1 Interconnected pharmaceuticals companies
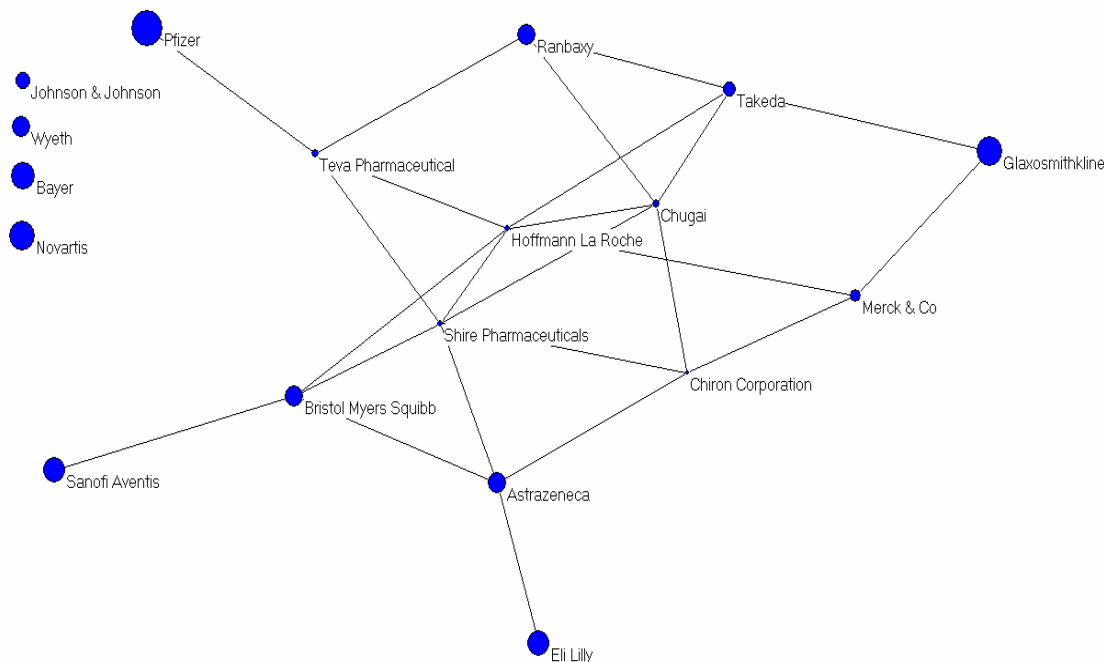
In this network map Pfizer is the most strongly connected to Teva Pharmaceutical, or the two firms appear strongly connected to each other – being discussed and referred to in the same context of URL-pages. AstraZeneca and Bristol Meyers Squibb are more central – discussed in the context of other firms such as: Sanofi Aventis, Hoffmann La Roche, Eli Lilly, Shire Pharmaceuticals, Chiron Corporation. The most central in this network map are: Hoffmann La Roche, Shire Pharmaceuticals, and Chugai, or a set of fairly small players that are discussed mainly in the context of all other firms. The four isolates in this map: Novartis, Johnson & Johnson, Wyeth, and Bayer are equally connected to all pharmaceutical companies, so their impact in the field is in the context of all other competitors.

Managerial Implications and Future Research

1. The volume of URL-pages that discuss issues related to European Pharmaceuticals and European healthcare has grown significantly – particularly during the first 7 months of this year. This is a dynamic public space where new stories appear continuously, shifting the attention to specific issues. Monitoring of this space is essential in order **to track major shifts in public opinion.**

2. There are two major players in the blog-space. One is generalist news, where information on healthcare and pharmaceutical issues appears. The other is specialist medical and healthcare

news with more in-depth information on disease areas and methods of treatment. Both players attract fairly similar public attention in terms of comments and interactions, which is still very low (within 1-2 days only after the publication). The majority of blogs have some association with private organisations and interests that manage the blogs, which suggests that serious and long-lasting blogs will exhibit the influence of some organisational agendas. Further research into specific blogs is essential in order to **monitor the evolution of these blogs.**

3. At present, most blogs are registered with a copyright disclaimer, which does not allow direct research of content beyond observations. If permissions are obtained from copyright holders, an **in-depth analysis of the text and the structure of these blogs** can reveal what concepts, ideas and values span the boundaries between comments to a post, between posts, and between threads across pages. These findings can be interpreted as a network structure of knowledge and opinions in the blog-space, and can identify strong semantic links in the public discussion.

4. Due to the high volume of entries in the blog-space, research is recommended on a narrow set of categories to demarcate a narrow semantic space for blog search and for analysis. Our choice of the 19 pharmaceutical companies is a successful strategy, as it can draw clear boundaries for the population of URL-pages in the database. **A repetition of this search strategy is recommended at short intervals** – in order to map changes in the blog-space, and to identify emerging trends.

5. The four semantic blocks that were identified in our search (DRUGS, DISEASE, INDUSTRY, and INSTITUTIONS) require independent in-depth research, as the current results are produced with the database that aimed at a complete representation of pharmaceutical firms. **Representative research on each semantic blog** will reveal in-depth associations, meaning and values that underpin discussions in each semantic blog. Such results will have a direct use in marketing and public relations.

6. Some of the technical outcomes from this project can be subjected to further in-depth analysis for **knowledge management** purposes. For example, the results in Appendix 3 show how the name of Pfizer is associated with all other pharmaceutical companies. **In-depth analysis of these associations** will reveal findings that will have implications for competition strategies.

7. The unique methodology that we used enables us to retrieve information on blogs that enable **blog ranking according to their importance in a selected semantic field** (rather then indiscriminate ranking by search engines). The sample of URL-pages with the highest ranking in Appendix 4 reveals a set of key actors in the blog-space with potentially different impact. Further monitoring of these blogs is essential in order to track their evolution and impact.

Appendix 1: Key-words for selection criteria

health	drug	industry	disease	regulation
European	drug development	pharmaceutical	diabetes	compulsory license
healthcare	drug treatment	industry	alzheimers	litigation
public health	drug efficacy	pharmaceutical	sex health	adverse event
health care	drug pricing	European	infectious	risk
health service	prescription drug	pharmaceutical	viruse	trust
health care service	generic drug	global	virology	International
health care system	over the counter	pharmaceutical	tropical	conference on
health policy	drug	biotech	oncology	harmonization
health technology	counterfeit drug	drug companies	cancer	Center for Drug
assessment	biomarker	drug company	blood pressure	Evaluation and
health tourism	biologics	**Pfizer**	lipid	Research
medical tourism	drug safety	**Glaxosmithkline**	cholesterol	European
medicine	drug trial	**Sanofi Aventis**	urology	Medicines
healthcare business	drug testing	**Novartis**	gastro	Evaluation Agency
healthcare trust	clinical trial	**Hoffmann La**	gastrointestinal	Food and Drug
hospital	adverse effect	**Roche**	gastric ulcer	Administration
private hospital	side effect	**Astrazeneca**	intestinal	Medicines and
in-patients	product recall	**Johnson &**	neuroscience	Healthcare
out patients	reimbursement	**Johnson**	central nervous	Products
patients	pharmacy	**Merck & Co**	system	Regulatory Agency
diagnostic		**Wyeth**	metabolic	National Institute
		Eli Lilly	metabolism	for Health and
		Bayer	allergy	Clinical Excellence
		Lacer	respiratory	community
		Bristol Myers		communities
		Squibb		charities
		Shire		charity
		Pharmaceuticals		
		Chiron		**region**
		Corporation		Europe
		Chugai		France
		Takeda		Spain
		Teva		Germany
		Pharmaceutical		United Kingdom
		Ranbaxy		England

Appendix 2: Blog population in database (Table 1b)

Key word	Total Google search results	The most relevant results	- less duplicates	- less 'shell-pages' = total in final database	% of pages in interconnected core	% of unique pages (periphery)
Pfizer	2363	350	296	**279**	69%	31%
GlaxoSmithKline	2408	151	149	**205**	76%	24%
Novartis	614	263	263	**194**	69%	31%
Bayer	1259	252	252	**159**	52%	48%
Sanofi Aventis	324	273	200	**156**	65%	35%
Eli Lilly	382	286	213	**143**	74%	26%
Ranbaxy	132	101	101	**106**	61%	39%
Wyeth	2136	142	142	**101**	63%	37%
AstraZneca	285	111	111	**100**	84%	16%
Bristol Myers Squibb	262	120	120	**94**	83%	17%
Johnson & Johnson	413	169	169	**72**	51%	49%
Takeda	61	57	57	**56**	89%	11%
Merck & Co	142	391	378	**39**	72%	28%
Chugai	26	23	23	**18**	50%	50%%
Teva Pharmaceuticals	22	20	20	**17**	82%	18%%
Hoffmann-La Roche	33	16	16	**11**	73%	27%
Shire Pharmaceuticals	10	9	9	**7**	86%	14%
Chiron Corporation	14	13	13	**6**	67%	33%
Lacer	6	6	6	**0**		
European Pharmaceutical	845	171	170	**83**	42%	58%
European healthcare	87	71	70	**45**	16%	84%
Total URL-pages	**11824**	**2995**	**2778**	**990**	**36%**	**64%**

Appendix 3: Co-reference of firms in blogs

	Pfizer	Glaxosmithkline	Sanofi Aventis	Novartis	Hoffmann La Roche	Astrazeneca	Johnson & Johnson	Merck & Co	Wyeth	Eli Lilly	Bayer	Bristol Myers Squibb	Shire Pharmaceuticals	Chiron Corporation	Chugai	Takeda	Teva Pharmaceutical	Ranbaxy	European healthcare	European pharmaceutical	Total
Pfizer	**279**	86	53	84	5	47	25	16	41	66	44	49	6	1	5	32	11	40	3	24	**279**
Glaxosmithkline	86	**205**	52	68	6	47	17	16	36	60	39	50	4	2	4	35	4	30	2	9	**205**
Sanofi Aventis	53	52	**156**	44	2	35	16	10	30	33	26	37	3	1	1	15	4	20	2	11	**156**
Novartis	84	68	44	**194**	2	41	14	7	29	39	37	31	4	1	6	24	6	36	4	10	**194**
Hoffmann La Roche	5	6	2	2	**11**	3	1	2	1	3	3	4	1	1		3	1	2			**11**
Astrazeneca	47	47	35	41	3	**100**	8	3	22	40	28	33	5	2	2	16	3	12		9	**100**
Johnson & Johnson	25	17	16	14	1	8	**72**	3	9	14	6	9				1	1	5		2	**72**
Merck & Co	16	16	10	7	2	3	3	**39**	4	5	2	7		3		1	1	2		2	**39**
Wyeth	41	36	30	29	1	22	9	4	**101**	28	25	21	3	1	2	15	3	16		2	**101**
Eli Lilly	66	60	33	39	3	40	14	5	28	**143**	35	34	4	1	3	23	1	23		10	**143**
Bayer	44	39	26	37	3	28	6	2	25	35	**159**	28	2	2	2	20	1	12	2	8	**159**
Bristol Myers Squibb	49	50	37	31	4	33	9	7	21	34	28	**94**	4			16	3	16		8	**94**
Shire Pharmaceuticals	6	4	3	4	1	5			3	4	2	4	**7**	1	1	2	1	2		3	**7**
Chiron Corporation	1	2	1	1		2		3	1	1	1	1	1	**6**	1	1		1		1	**6**
Chugai	5	4	1	6	1	2			2	3	2		1	1	**18**	6		6		2	**18**
Takeda	32	35	15	24	3	16	1	1	15	23	20	16	2	1	6	**56**	3	19		4	**56**
Teva Pharmaceutical	11	4	4	6	1	3	1	1	3	1	1	3	1			3	**17**	8			**17**
Ranbaxy	40	30	20	36	2	12	5	2	16	23	12	16	2	1	6	19	8	**106**	1	3	**106**
European healthcare	3	2	2	4							2							1	**45**		**45**
European pharmaceutical	24	9	11	10		9	2	2	2	10	8	8	3	1	2	4		3		**83**	**83**
Total	**279**	**205**	**156**	**194**	**11**	**100**	**72**	**39**	**101**	**143**	**159**	**94**	**7**	**6**	**18**	**56**	**17**	**106**	**45**	**83**	**990**

Appendix 4: Ranking of the top-blogs

Rank_size	Rank_links	Rank_keyword	Rank_URL_from_blog	Lable_Numer	PageTitle	PageURL	Month	Year	Blog_name	Blog_link
12		1	5	242.3	New Evidence Shows MabThera Inhibi	http://google-sin	Jul	2,007	google-sina Medical,health a	http://google-sina.com
	41	4		379.1	Weekly Medical News Newsletter - 25	http://medicalne	Jun	2,007	Medical News	http://medicalnewsreview.
		7	3	205.2	Contents Click on content link below to	http://fadetheblc	Mar	2,006	Fade the Blog	http://fadetheblog2.blogsp
		8		786	Welcome to Rx Blog	http://www.rxblo	Sep	2,006	RxBlog (tm) .net / Health + [	http://www.rxblog.net/
28		9		350	25 new messages in 18 topics - digest	http://lib.dropin.	Jul	2,007	Talk: Med	http://lib.dropin.org/med/
		13		128	Canada's Shame	http://canadas-s	Jul	2,007	Canada's Shame	http://canadas-shame.blo
		14		234	Sobre activismo e Boehringer	http://gatportuga	Apr	2,006	O blog do GAT	http://gatportugal.blogspot
34		15		209	New Oral Agents for Multiple Sclerosis	http://feeds.rapi	Mar	2,007	MSRC Latest MS News	http://feeds.rapidfeeds.cor
5		17		339	Faculty Index	http://laparosco	Sep	2,006	SLS Annual Meeting and En	http://laparoscopy.blogs.c
		18		109	Industrial Recruiting Program in Chemi	http://brevetoxin	Apr	2,007	March Loves Chemistry	http://brevetoxin.spaces.li
		30	4	675.4	SA not doing so hot....hmmmm	http://www.cafe	Aug	2,007	Cafepharma Message Boar	http://www.cafepharma.co
		35	6	349.2	25 new messages in 13 topics - digest	http://lib.dropin.	Jul	2,007	health.alternative & fitness.w	http://lib.dropin.org/health/
		43	6	744.3	Bayer HealthCare And Regeneron Initi	http://www.medi	Aug	2,007	Medical News and Health N	http://www.medicalnewsbl
		48	1	382.1	INEGY? (ezetimibe/simvastatin) Is Sup	http://medicine.t	Jan	2,007	Medicine, drugs and health	http://medicine.truckofnew
7	1			354	BlogsCNN Exchange Notebook BlogA	http://listrss.blog	Jun	2,007	listrss	http://listrss.blogspot.com/
22	3			395	american history picture x send free rin	http://middle-iro	Oct	2,006	middle-irocic	http://middle-irocic.blogsp
2	6		4	467.2	[TrackEngine] Unknown News	"News	http://pbgiltner.t	May	2,007	Giltner Review	http://pbgiltner.blogspot.cc
30	11			117	European Pharmaceutical Marketing R	http://business-i	Jul	2,007	Business Internet Live	http://business-internet-liv
15	12			141	RSS 2 E-MAIL: Clipmarks	Live Clips	http://clipmark.b	May	2,007	Clipmarks	http://clipmark.blogspot.cc
	13			28	ketchikan payday loan	http://alexiaqphi	Aug	2,007	Alexia Q Phipps Blog	http://alexiaqphipps2000.t
	14			510.1	brothers birthday greeting cards	http://reillyzpratt	Aug	2,007	Reilly Z Prather Blog	http://reillyzprather2005.bl
	15			32	moodys financial software	http://almaswhe	Aug	2,007	Alma S Wheeler Blog	http://almaswheeler1997.t
	18			323	online translation language	http://joelscahill:	Aug	2,007	Joel S Cahill Blog	http://joelscahill2002.blogs
	19		6	282.3	The Euro-Next Biotech Bubble?	http://hotnewz.ir	Jul	2,007	The Best News Portal	http://hotnewz.info
18	24			642.2	Another Canadian Speaks Against Hea	http://wizbangbl	Jul	2,007	Wizbang	http://wizbangblog.com/
	26			670	Before the bell: Friday's plunge points t	http://www.blog	Aug	2,007	bloggingstocks	http://www.bloggingstocks
9	30			107.2	Hocks_On-Line_Pharmacy XXVII	http://brandpile.l	Aug	2,007	Brand Pile Weblog - Blog	http://brandpile.blogspot.c
1	38			311	Fight Terrorism: Legalize Heroin	http://intermexfr	Aug	2,007	INTERMEX POWER	http://intermexfreemarket.
10				327	Evidence-Based Guidelines for Cardio	http://josenader:	May	2,007	Transicion menopausica	http://josenader2007.blogs
21				662.1	Week #46 (2006-11-18) in Review (FIN	http://www.billca	Nov	2,006	Bill Cara	http://www.billcara.com/
				159	sviki.doc	http://davidukral	Apr	2,007	Mocjicjemo	http://davidukralju.blogspc
32			2	341.4	31/07 SoloStar/Gerresheimer/risk man	http://latestpubli	Jul	2,007	Latest news release	http://latestpublication.blo
			6	197.1	Medical news - Group Psychotherapy (	http://engineersi	Jul	2,007	Medical News Updates	http://engineersindia.org/h
			4	770.5	12-JUL-07: YAZ approved in Netherlan	http://www.phar	Jul	2,007	Pharmaceutical Executive E	http://www.pharmexeceur
			5	766.3	FDA's Tighter Approval Measures Affe	http://www.phar	Aug	2,007	Pharmaceutical Internationa	http://www.pharmaceutica
			5	476.4	GSK - Avandia: Doug Arbesfeld, come	http://pharmago	Aug	2,007	pharmagossip	http://pharmagossip.blogs
			6	768.2	Sanofi-Aventis To Limit UK Wholesaler	http://www.phar	Jul	2,007	Pharmalot	http://www.pharmalot.com
			4	769.2	Weekly Roundup: July 23, 2007	http://www.phar	Jul	2,007	pharmasia News	http://www.pharmasianew