STRATEGIC INNOVATION MANAGEMENT ON THE BASIS OF SEARCHING AND MINING PRESS RELEASES

Jan Finzen, Maximilien Kintz, Holger Kett
Fraunhofer IAO, Nobelstr. 12, 70195 Stuttgart, Germany

Steffen Koch
Visualization and Interactive Systems Group (VIS), Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany

Keywords: Web Information Retrieval, Press Releases, Strategic Innovation Management.

Abstract: Press releases may contain a lot of information that is especially applicable in strategic innovation management: They contain up-to-date information by definition and thus may give hints to upcoming trends and techniques. They also tell a lot about the strategies of partners, customers, and, most of all, competitors. We analysed many of today's existing press release search engines and identified a number of shortcomings: The query frontends do not provide enough flexibility with regards to search space restriction, the result lists presentation typically cannot be influenced by the user, and the ranking order stays often unclear. Press releases offer a number of features that make them useful for automatic handling but are widely ignored by today's search engines: They are relatively homogenously structured and contain certain kinds of easy-extractable meta-data that can be utilized for use cases such as monitoring trends (date of publishing), discovering geographical competency clusters (author and address information), or identifying relations between companies (firm name mentioning). We describe the prototype of a new press release search engine that makes use of the above-mentioned meta-data and additionally offers sophisticated search features tailored to the needs of innovation professionals.[1]

1 INTRODUCTION

Strategic innovation management deals with "the planning, organisation, execution, and control of any innovation activity aiming at the creation and preservation of competitive advantages" (Goos and Hagenhoff, 2003). A lot of an innovation manager's daily work is somehow related to searching the web for information. A recent survey among industrial innovation actors and experts states that web search engines clearly dominate the internet-based technologies used by innovation actors, followed by online editions of professional journals (cf. Novanet, 2006).

Some types of documents, accessible online, are of special value for strategic innovation management because they are very likely to include innovation-relevant information.

Press releases represent such a document type. They are highly relevant within strategic innovation management, as they can be expected to contain some information that is "new" in one way or another (in fact the term "news release" is often used as a synonym). Press releases are usually authored by organisations to spread some information considered worth publishing and constitute a means for public relations. They target media editors who are expected to pick up the topic and develop articles on it.

Press releases can tell a lot about activities and plans of customers, partners, as well as competitors and may thus be applied as a source of information for business or competitive intelligence which we see as sub area of strategic innovation management.

As of today, most press release portals offer rather limited search functions and largely ignore the intrinsic features of press releases such as location and date information.

Within this paper we introduce a prototype of a sophisticated press release search engine that takes into account a number of innovation professionals' requirements. The remainder of this paper is structured as follows:

Finzen J., Kintz M., Kett H. and Koch S.
STRATEGIC INNOVATION MANAGEMENT ON THE BASIS OF SEARCHING AND MINING PRESS RELEASES.
DOI: 10.5220/0001840803470353
In *Proceedings of the Fifth International Conference on Web Information Systems and Technologies (WEBIST 2009)*, page
ISBN: 978-989-8111-81-4

In Section 2, we analyse the state of the art regarding press release search engines. We identify several shortcomings and propose methods to overcome them within a meta search engine approach based on RSS data feed integration.

In Section 3, we introduce the current state of our search engine's prototype and explain the most important features using real world examples.

In Section 4 we discuss the results. We point out, what problems still have to be overcome and give an outlook to future activities.

2 METHODS AND MATERIALS

In this section, we first define the specific information seeking ambitions of "innovation officers". Keeping this in mind, we analyse the state of the art regarding press release search engines. We identify several shortcomings and propose methods to overcome them by means of a meta search engine approach based on RSS data feed integration.

2.1 Target Group Analysis

There is no official definition of an "innovation professional". Within the context of our work, we define the strategic innovation professional as a specialist who is taking care of market development and oversees the actions of customers, partners, and competing companies. Regarding the innovation officer's information retrieval abilities, we account him to be rather motivated in investing considerable effort to retrieve high-quality results from search tasks. Stock and Lewandowski (2006) distinguish three different classes of search engine users:

1. Information professionals: Experts for searching and finding documents with professional background in information sciences. They know relevant databases, develop search strategies and use professionally elaborated queries. Research is predominantly done by order of a third party.
2. Professional end users: Business experts. Searching because of their business-related information needs. They know the suitable business-specific data bases and search engines, but rarely develop research strategies and professionally elaborated queries.
3. Amateurish users: Searching mostly for private and sometimes for professional reasons. Restricting their efforts to query the surface web using a general-purpose search engine (e.g. Google). They do not apply systematic search strategy and normally use short queries.

Innovation professionals do not match any of those categories exactly but rather combine attributes of the first two classes: They are experts concerning their companies' business. But they are expected to provide a wider view and to be able to "think outside the box". Therefore, they bring along a broad knowledge of where and how to find information meeting their requirements and needs. Innovation professionals perform searches in a frequent and strategic manner – thus, they know how to effectively use search engines.

2.2 State of the Art Analysis

Today, press releases are most often disseminated using commercial press release distribution services.

Based on web research and experts' opinions we identified the most important German press release portals and some of the largest US portals relevant for our examination. We analysed these portals with regard to the following aspects:

- *Business model*: How do they earn money?
- *Amount of content*: How many press releases are published?
- *Search frontend features*: How easily and detailed can a query be specified?
- *Search result display*: How are search results presented to the user?

2.2.1 Business Model

Usually the business models of these services are based on advertisement (almost all the services we know use Google AdSense). In some cases, the author of a press release has to pay for the publishing. In these cases, the charges amount to an average of 100 to 300 Euros per release, depending on the diffusion rate and additional services like click-tracking etc.

2.2.2 Amount of Content

The amount of content, i.e., the publishing activity of the surveyed press release distribution portals varies significantly. While only a few portals publish up to 800 press releases on average each day, most portals restrict themselves to less than 50 releases each day. Thus, the distribution curve forms a typical long-tail (cf. figure 1).

Table 1: Evaluation results of examined press release websites (excerpt).

Name	Language	Publishing costs (€)	Query frontend quality	Result presentation quality	RSS feed quality	Total query quality	Press releases per day	Amount of content
PR Newswire	EN	120	●○	●○	●●●	●●	800	●●●
PresseEcho	DE	-	○	●	●	●	500	●●●
PR Web	EN	80-230	●	○	○	○	500	●●●
OpenPR	DE	-	○	●	○	○	280	●●●
PressePortal	DE	315	●	●●	●●●	●●	200	●●●
INAR	DE	-	○	○	●●	●	175	●●
LifePR	DE	99	●●	●●	●●	●●	130	●●
PM-Webservice	DE	-	●○	●	●	●	80	●
Firmenpresse	DE	-	●	●	●●	●	70	●
Online Artikel	DE	-	●	●○	●	●	50	●
DailyNet	DE	-	●	●	●	●	50	●
Pressbot	DE	-	●	○	●	●	50	●
AlphaGalileo	EN-DE	50-100	●	●●	●●	●●	25	○
PresseText	DE	167	●●	●○	○	●	20	○
OpenBroadcast	DE	-	●	●	○	●	20	○
PressText	DE	160	●●	●●	-	●	10	○
Press1	DE	-	○	●	●●	●	5	○

Explanation of columns: Publishing costs per press release; Query frontend quality: ●●● = many options and easy to use, ○ = few options, hard to use; Result presentation quality: ●●● = very clear and complete presentation, ○ = few information, unclear presentation; RSS feed quality: ●●● = feed complete, ○ = important data missing, - = no feed available; Total query quality: mean of query frontend, result presentation, and RSS feed quality; Amount of content: ●●● = 200 and above, ●● 100 to 200, ● = 50 to 100, ○ = 0 to 50.

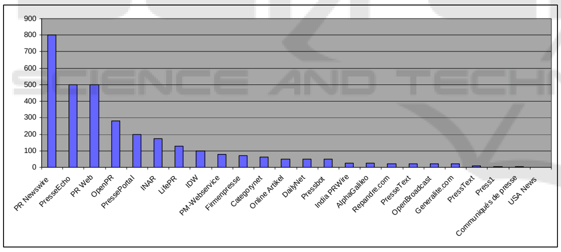


Figure 1: Long-tail distribution of amount of content.

2.2.3 Search Frontend Features

All the web sites we looked at offered basic search functionality that allows searching for key words in the full text of all press releases. Many sites additionally offer an “expert search” that allows further restriction of the search space by specifying one or more of the following constraints:

- The publishing time frame
- Boolean combinations of key words,
- Geographical regions (e.g., countries or states)
- Company names, or
- Categories (whereby the granularity of classification systems varies significantly).

Remarkably, none of the web sites offer all of those features. However, we consider all of these constraints useful, especially if the user repeatedly searches with professional background, as it is the case within the strategic innovation management.

2.2.4 Search Result Presentation

While all portals present their search results as lists which link to the complete text, as known by standard search engines, the information displayed in the list varies significantly, too: For example, PRNewswire restricts itself to the title, the author’s name and the date of publishing. LifePR also displays the company name, category information, the first sentences of the text and, if available, also images. The number of results per page ranges from 10 to all. But in most cases this cannot be configured by the user. The ranking algorithm often remains unclear and can only rarely be influenced by the user.

2.3 Examination Results

Table 1 shows an excerpt of our examination results ordered by the amount of content. As many different press release portals exist and press releases are often exclusively published by one of them, a meta search approach seems promising to increase the recall and improve the ease of use. Both the query frontend and the presentation of search results offer room for improvement. Additional usage of meta data for information aggregation, for trend monitoring, or identification of competency clusters is not provided at all in today's press release portals. However, these approaches seem quite promising for innovation management, especially when they are applied in combination with a meta search engine approach.

3 AN IMPROVED PRESS RELEASE META SEARCH ENGINE

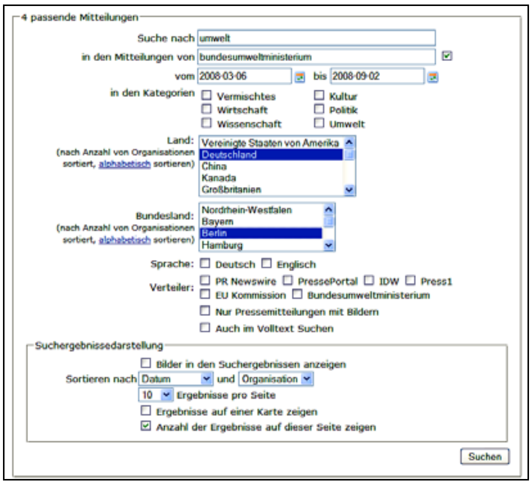
Based on our findings we designed and implemented a press release meta search engine which,

- is based on RSS feed aggregation,
- follows a best-of-breed approach with regards to the query and result frontend, and
- offers additional search features based on meta data like dates and addresses.

In the following we will give an overview of the prototype's current state and illustrate our search and mining approaches using real-world examples.

3.1 Query Frontend

As we stated in the previous section, the query frontends of today's press release search engines do lack the options for restricting the search space according to the user's needs. Thus, we defined an expert query frontend that allows an arbitrary combination of all the search space constraints listed in Section 2.2.3 (see Figure 2).



4 passende Mitteilungen

Suche nach: umwelt

in den Mitteilungen von: bundesumweltministerium ☑

vom 2008-03-06 bis 2008-09-02

in den Kategorien: □ Vermischtes □ Kultur □ Wirtschaft □ Politik □ Wissenschaft □ Umwelt

Land: (nach Anzahl von Organisationen sortiert, alphabetisch sortieren)
Vereinigte Staaten von Amerika
Deutschland
China
Kanada
Großbritanien

Bundesland: (nach Anzahl von Organisationen sortiert, alphabetisch sortieren)
Nordrhein-Westfalen
Bayern
Berlin
Hamburg

Sprache: □ Deutsch □ Englisch

Verteiler: □ PR Newswire □ PressePortal □ IDW □ Press1 □ EU Kommission □ Bundesumweltministerium

□ Nur Pressemitteilungen mit Bildern

□ Auch im Volltext Suchen

Suchergebnissedarstellung

□ Bilder in den Suchergebnissen anzeigen

Sortieren nach Datum und Organisation

10 Ergebnisse pro Seite

□ Ergebnisse auf einer Karte zeigen

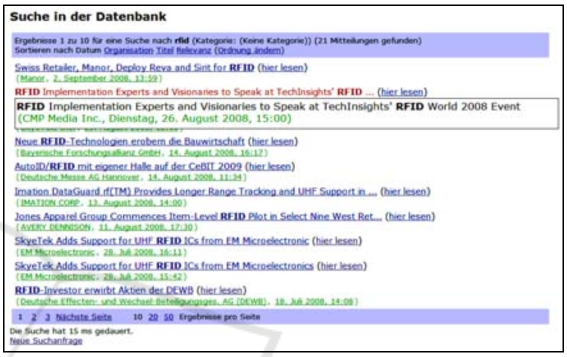
☑ Anzahl der Ergebnisse auf dieser Seite zeigen

Figure 2: Configuring the search space.

This expert query frontend is intended to be used by motivated experienced users. In addition, simple ad hoc queries can be entered using a query dialog with reduced complexity.

3.2 Search Result Presentation

We followed a "best-of-breed"-approach when designing the result list presentation. Unlike the press release search engines we analysed, our new prototype lets the user influence both the number of results displayed per page and the ranking of the list (see Figure 3).



Suche in der Datenbank

Ergebnisse 1 zu 10 für eine Suche nach **rfid** (Kategorie: (Keine Kategorie)) (21 Mitteilungen gefunden)
Sortieren nach Datum Organisation Titel Relevanz (Ordnung ändern)

Swiss Retailer, Manor, Deploy Reva and Sirit for **RFID** (hier lesen)
(Manor, 2. September 2008, 13:59)

RFID Implementation Experts and Visionaries to Speak at TechInsights' **RFID** ... (hier lesen)

RFID Implementation Experts and Visionaries to Speak at TechInsights' **RFID** World 2008 Event
(CMP Media Inc., Dienstag, 26. August 2008, 15:00)

Neue **RFID**-Technologien erobern die Bauwirtschaft (hier lesen)
(Bayerische Forschungsallianz GmbH, 14. August 2008, 16:17)

AutoID/**RFID** mit eigener Halle auf der CeBIT 2009 (hier lesen)
(Deutsche Messe AG Hannover, 14. August 2008, 11:34)

Imation DataGuard rf(TM) Provides Longer Range Tracking and UHF Support in ... (hier lesen)
(IMATION CORP, 13. August 2008, 14:00)

Jones Apparel Group Commences Item-Level **RFID** Pilot in Select Nine West Ret... (hier lesen)
(AVERY DENNISON, 11. August 2008, 17:30)

SkyeTek Adds Support for UHF **RFID** ICs from EM Microelectronic (hier lesen)
(EM Microelectronic, 28. Juli 2008, 16:11)

SkyeTek Adds Support for UHF **RFID** ICs from EM Microelectronics (hier lesen)
(EM Microelectronic, 28. Juli 2008, 15:42)

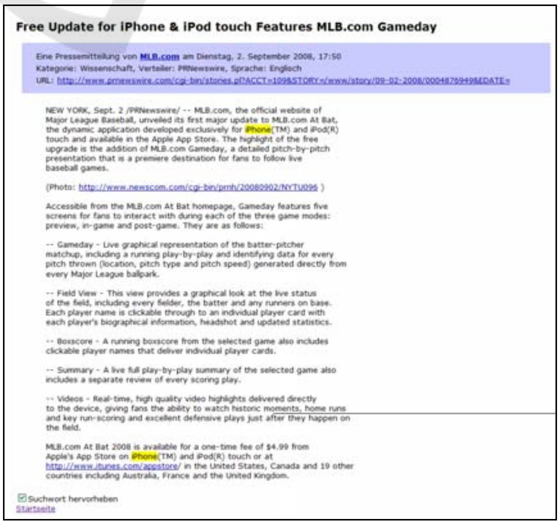
RFID-Investor erwirbt Aktien der DEWB (hier lesen)
(Deutsche Effecten- und Wechsel-Beteiligungsges. AG (DEWB), 18. Juli 2008, 14:08)

1 2 3 Nächste Seite 10 20 50 Ergebnisse pro Seite

Die Suche hat 15 ms gedauert.
Neue Suchanfrage

Figure 3: Search results presentation.

As we follow a meta search engine approach the found results may well come from different press release distribution sites thus looking quite inhomogenously when displayed on the original sites. To improve readability we added a uniform result display that shows each result in a similar look and feel (see Figure 4).



Free Update for iPhone & iPod touch Features MLB.com Gameday

Eine Pressemitteilung von MLB.com am Dienstag, 2. September 2008, 17:50
Kategorie: Wissenschaft, Verteiler: PRNewswire, Sprache: Englisch
URL: http://www.prnewswire.com/cgi-bin/stories.pl?ACCT=109&STORY=/www/story/09-02-2008/0004876949&EDATE=

NEW YORK, Sept. 2 /PRNewswire/ -- MLB.com, the official website of Major League Baseball, unveiled its first major update to MLB.com At Bat, the dynamic application developed exclusively for iPhone(TM) and iPod(R) touch and available in the Apple App Store. The highlight of the free upgrade is the addition of MLB.com Gameday, a detailed pitch-by-pitch presentation that is a premiere destination for fans to follow live baseball games.

(Photo: http://www.newscom.com/cgi-bin/prnh/20080902/NYTU096)

Accessible from the MLB.com At Bat homepage, Gameday features five screens for fans to interact with during each of the three game modes: preview, in-game and post-game. They are as follows:

-- Gameday - Live graphical representation of the batter-pitcher matchup, including a running play-by-play and identifying data for every pitch thrown (location, pitch type and pitch speed) generated directly from every Major League ballpark.

-- Field View - This view provides a graphical look at the live status of the field, including every fielder, the batter and any runners on base. Each player name is clickable through to an individual player card with each player's biographical information, headshot and updated statistics.

-- Boxscore - A running boxscore from the selected game also includes clickable player names that deliver individual player cards.

-- Summary - A live full play-by-play summary of the selected game also includes a separate review of every scoring play.

-- Videos - Real-time, high quality video highlights delivered directly to the device, giving fans the ability to watch historic moments, home runs and key run-scoring and excellent defensive plays just after they happen on the field.

MLB.com At Bat 2008 is available for a one-time fee of $4.99 from Apple's App Store on iPhone(TM) and iPod(R) touch or at http://www.itunes.com/appstore/ in the United States, Canada and 19 other countries including Australia, France and the United Kingdom.

☑ Suchwort hervorheben
Startseite

Figure 4: Press release presentation.

Since innovation officers carry out long-ranging search tasks quite regularly, search queries need to be stored and automatically repeatedly carried out within planned intervals. In our search engine, the user may subscribe to any query she likes and obtain the search results as RSS data feeds, again.

3.3 Meta Data Usage

Press releases, apart from the actual content, contain additional information that can be utilised within the search process.

- *Title*: Can be used for duplicate detection, which is quite important within a meta search engine approach
- *Abstract*: If it exists, it can be displayed within the result list to improve usability and supersedes algorithmic summarization approaches.
- *Date*: Allows restricting the search space and time-tracking of press releases.
- *Author and address information*: Apart from restricting the search space, this information can be utilised to group releases, e.g. by company. Address information can be geo-coded and allow sophisticated geographical views.
- *Tags and categories*: Allow grouping of search results.

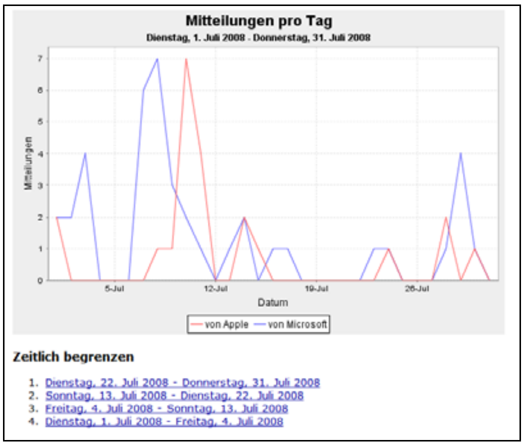
In the following we will introduce how our prototype, which exploits this information to offer advanced search features, is tailored to the needs of strategic innovation management.

3.3.1 Monitoring Trends and Detecting Events

The date of publishing is included in every press release. Thus, it is quite easy to keep track of the number of results for an arbitrary search query and visualize the results as diagrams.

In Figure 5 we illustrate this concept by contrasting the number of press releases of two different companies. The red line indicates the number of press releases published by apple between July 1st and July 31st. Note the peek on July 10th which can be traced back to the fact that the iPhone G3 was released on this very day. Similar peeks were observed when the new iPod generation was introduced in September 2008.

If combined with a keyword-based search for an arbitrary topic, trends can be monitored. Figure 6 shows the temporal distribution for the keyword "Finanzkrise" (financial crisis) between July and November 2008.



Zeitlich begrenzen

1. Dienstag, 22. Juli 2008 - Donnerstag, 31. Juli 2008
2. Sonntag, 13. Juli 2008 - Dienstag, 22. Juli 2008
3. Freitag, 4. Juli 2008 - Sonntag, 13. Juli 2008
4. Dienstag, 1. Juli 2008 - Freitag, 4. Juli 2008

Figure 5: Time-tracking press releases of companies.

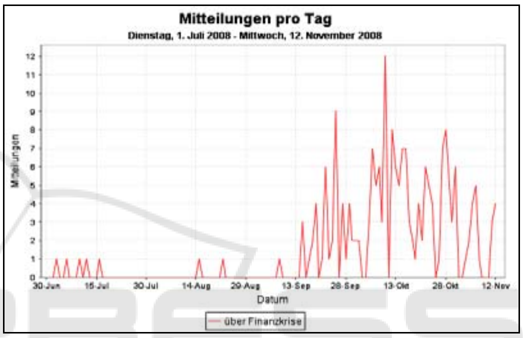
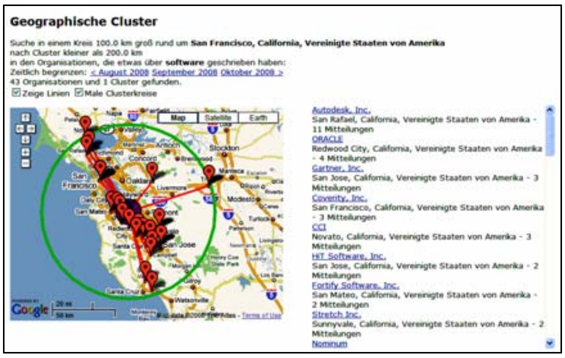


Figure 6: Time-tracking the topic "financial crisis".

We complemented the line chart view with a bar chart view that visualises shorter time-frames better. If the user clicks on the diagram the search results for the underlying query restricted to the corresponding time-frame are displayed as a list (see Figure 3).

3.3.2 Identifying Geographical Competence Clusters

Similar to the application of dates for visualising the temporal distribution of press releases we can use geographical information included in press releases to depict geographical distributions. This feature allows straight-forward identification of competence clusters by showing them on maps. Figure 7 shows a simple example: If we search for "software"-related press releases we can easily identify a high publication density south of the San Francisco bay area which of course is due to many software companies being located around the Silicon Valley. For our example we combined the Google Maps mashup service with a Quality Threshold Clustering (Heyer et al, 1999) – based algorithm to detect an arbitrary number of clusters.



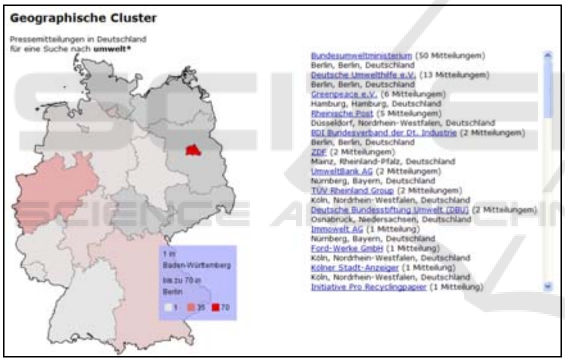
Geographische Cluster

Suche in einem Kreis 100.0 km groß rund um **San Francisco, California, Vereinigte Staaten von Amerika**
nach Cluster kleiner als 200.0 km
in den Organisationen, die etwas über **software** geschrieben haben:
Zeitlich begrenzen: < August 2008 September 2008 Oktober 2008 >
43 Organisationen und 1 Cluster gefunden.
☑ Zeige Linien ☑ Male Clusterkreise

Autodesk, Inc.
San Rafael, California, Vereinigte Staaten von Amerika - 11 Mitteilungen
ORACLE
Redwood City, California, Vereinigte Staaten von Amerika - 4 Mitteilungen
Gartner, Inc.
San Jose, California, Vereinigte Staaten von Amerika - 3 Mitteilungen
Coverity, Inc.
San Francisco, California, Vereinigte Staaten von Amerika - 3 Mitteilungen
CCI
Novato, California, Vereinigte Staaten von Amerika - 3 Mitteilungen
HiT Software, Inc.
San Jose, California, Vereinigte Staaten von Amerika - 2 Mitteilungen
Fortify Software, Inc.
San Mateo, California, Vereinigte Staaten von Amerika - 2 Mitteilungen
Stretch Inc.
Sunnyvale, California, Vereinigte Staaten von Amerika - 2 Mitteilungen
Nominum

Figure 7: Cluster Visualisation.

Another way of visualising geographical distributions is colouring maps. This kind of visualisation is often used to highlight differences in distribution between geographical regions, e.g. election results. We implemented this view on the level of the German federal counties as well as of the US federal states. Figure 8 shows the distribution of press releases for the search term “umwelt” (environment) within the counties of Germany.



Geographische Cluster

Pressemitteilungen in Deutschland
für eine Suche nach **umwelt***

Bundesumweltministerium (50 Mitteilungem)
Berlin, Berlin, Deutschland
Deutsche Umwelthilfe e.V. (13 Mitteilungem)
Berlin, Berlin, Deutschland
Greenpeace e.V. (6 Mitteilungem)
Hamburg, Hamburg, Deutschland
Rheinische Post (5 Mitteilungem)
Düsseldorf, Nordrhein-Westfalen, Deutschland
BDI Bundesverband der Dt. Industrie (2 Mitteilungem)
Berlin, Berlin, Deutschland
ZDF (2 Mitteilungem)
Mainz, Rheinland-Pfalz, Deutschland
UmweltBank AG (2 Mitteilungem)
Nürnberg, Bayern, Deutschland
TÜV Rheinland Group (2 Mitteilungem)
Köln, Nordrhein-Westfalen, Deutschland
Deutsche Bundesstiftung Umwelt (DBU) (2 Mitteilungem)
Osnabrück, Niedersachsen, Deutschland
Immowelt AG (1 Mitteilung)
Nürnberg, Bayern, Deutschland
Ford-Werke GmbH (1 Mitteilung)
Köln, Nordrhein-Westfalen, Deutschland
Kölner Stadt-Anzeiger (1 Mitteilung)
Köln, Nordrhein-Westfalen, Deutschland
Initiative Pro Recyclingpapier (1 Mitteilung)

Figure 8: Colouring Maps.

3.3.3 Identifying Relations between Companies

Furthermore, we implemented an algorithm to identify relations between two companies. For each company that publishes a press release we find out:

- Which other companies have mentioned this company in their press releases and how often?
- Which other companies have been mentioned in press releases of this company and how often?

Within the current implementation status we only detect names of companies that have already published at least one press release themselves – and thus can be found in our data base. In future releases, a well-maintained list of company names will improve the approach.

Both - forward and backward citations - can be visualised as reference graphs. Figure 9 shows a reference graph for the company “Business Objects”. Citation frequency is encoded by the thickness of the corresponding edge. In figure 9 this can be observed by the fact that the edge between “Business Objects” and SAP is thicker, which correlates to the fact that Business Objects is a SAP company since the takeover in October 2007.

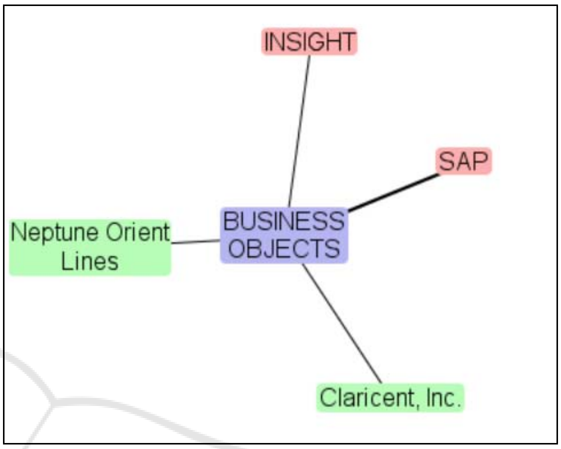


Figure 9: Company Reference Graph, Level 1.

The reference graph can be computed recursively – but time constraints have to be taken into account. The computation time highly depends on the number of different companies stored in the database and the number of referencing companies identified in the previous step. Within the current prototype we therefore restricted the reference graph computation at runtime to the second level (references of references). Nevertheless, the process can be accelerated using offline precomputation steps. As future improvement we will add heuristics for firm name unification (cf. Magnani and Montesi, 2007), in order to detect that for example, “Apple” and “Apple Comp. Inc” denote the same entity.

3.4 Implementation Notes

As nearly all of the analysed press release portals offer RSS based data streams, feed aggregation seems to be a promising means of integrating the different data sources within a meta search engine approach. On this basis, we started building a press release corpus by aggregating the newsfeeds of the five most active portals. Unfortunately, the provided RSS data streams do not include the full text but only a URL pointing to the HTML page. In some cases, meta information is also not explicitly marked-up. Thus, we built some additional page scrapers to extract both, the full text and the meta data from the website.

The current prototype consists of two applications:

- A standard Java application fetches the newest press releases every few minutes by reading the RSS feeds and web scraping additional data for any newly retrieved message.
- The search engine itself is implemented as a standard Java web application running on an Apache Tomcat web server.

All data is currently stored in a MySQL database, but we will soon migrate to a Lucene search index to improve retrieval performance.

4 DISCUSSION AND CONCLUSIONS

Within this paper we introduced a prototype of a new press release search engine that offers advanced search possibilities by taking into account the specific structure and meta data of press releases.

The findings so far indicate that systematic press release observation can help exposing facts and developments, which are highly useful within strategic innovation management. Nevertheless the examples we provided are still rather simple. To get more interesting results, more and more carefully selected data is needed.

One of our findings is that the publishing activity among the available press release distribution portals forms a typical long-tail. A way to improve the search data corpus size thus lies in integrating as many data sources as possible, i.e., not only the four or five biggest portals have to be accounted for, but also the many small providers.

The extraction and processing of meta data must be enhanced; for instance, company names like IBM and IBM Corp. are currently treated as two different organisations. In future releases we intend to apply named entity recognition algorithms to reliably identify and unify company names. Additionally, the usage of a well-maintained toponym reference list will enhance the geo locating functions. The classification of press releases currently depends on explicit markup, which is not offered by all providers. We are therefore evaluating a stochastic topic detection approach to automatically classify press releases based on their textual content.

To further improve the search result quality we are currently adding a semantic-based query expansion mechanism. This will increase the recall of the meta search engine and improve the outcomes of the trend monitoring and cluster identification approaches - as both rely heavily on the amount of relevant data retrieved.

The feedback we received from cooperating companies so far support our assumption that the tool is regarded useful by innovation professionals. But more systematic user studies have to be arranged to evaluate formally how useful the implemented features are with regard to different user types and different business areas.

As part of our further activities we will integrate the press release search engine as one module among others into a so-called innovation mining cockpit. This cockpit will form a single point of entry for all of the innovation professional's web search related activities (see Stathel et al, 2008, for details).

REFERENCES

Goss, P., and Hagenhoff, S., 2003. Strategisches Innovationsmanagement: Eine Bestandsaufnahme. In Schumann, M. (ed.), Arbeitsbericht Nr. 11/2003 des Instituts für Wirtschaftsinformatik der Georg-August-Universität Göttingen, Göttingen.

Heyer, L. J., Kruglyak, S., and Yooseph, S., 1999. Exploring Expression Data: Identification und Analysis of Coexpressed Genes. In *Genome Res.* 1999 9: 1106-1115.

Magnani, M., and Montesi, D., 2007. *A study on company name matching for database integration.* Technical Report UBLCS-07-15. May 2007.

Novanet, 2006. Information in der Internetökonomie. 2nd newsletter of the NovaNet project, http://www.nova-net.de/fhg/Images/nova-net_2-Newsletter_tcm231-60869.pdf. Accessed September 9th, .2008.

Stathel, S., Finzen, J., Riedl, C., and May, N., 2008. Service Innovation in Business Value Networks. In *Proceedings of the XVIII International RESER Conference.*

Stock, W. G., and Lewandowski, D., 2006. Suchmaschinen und wie sie genutzt werden. WISU 35(2006)8-9, 1078-1083.

[1] The project was funded by means of the German Federal Ministry of Economy and Technology under the promotional reference "01MQ07012". The authors take the responsibility for the contents.