

# Qualitative Corporate Dashboards for Corporate Monitoring

Peng Jia and Miklos A. Vasarhelyi<sup>1</sup>

## Introduction

Electronic Commerce<sup>2</sup> is accelerating dramatically changes in the business process. Electronic cataloging, item tracking, blind auctioning, and other innovations are fundamentally changing the way business is performed. Accounting, or the method that that business measures its performance, and Auditing, or the method that corporate stakeholders are assured of the accuracy of the reported results of a firm, are also going to undergo dramatic changes. The most dramatic of these are continuous reporting and auditing. In the current stage of progressive change, new needs and procedures are already emerging for corporate process monitoring and assurance. The AICPA, in its assurances services report<sup>3</sup>, has proposed a set of new assurance services including WebTrust, Systems Reliability, and Performance Measures. The AICPA jointly with the CICA<sup>4</sup> has issued its Continuous Auditing report arguing for **extensive** corporate process monitoring and the continuous audit of corporate systems. At the same time professional service firms are starting to evolve towards strategic auditing issues linking auditing methodology to strategic objective setting and corporate monitoring<sup>5</sup>. Two major changes in the auditing process per se will entail shorter period attestation (evolving towards continuous auditing, maybe opinion on demand), and the expansion of examined variables towards additional non-financial variables including human resources, production, customer care, costs, etc.

Kogan, Sudit and Vasarhelyi<sup>6</sup> have proposed a program of research on what they define as continuous online auditing (COA). This program is needed to develop a sound theoretical basis for the "audit of the future."

This paper expands the traditional financial audit framework arguing for the expansion of the scope of evidence collection to cover on-line corporate information (in particular news), using semantic analysis methods. Such an approach is aimed to extract valuable information from qualitative data and to ultimately present the analysis results on a media information dashboard with dynamic and continuous updating.

## The media dashboard

A media dashboard is designed to systematically present up to date news about a company through scrutiny of its media coverage. The media dashboard should not only selectively and systematically organize and present the company's on-line news, but also show the degree and nature that a company is reported in the media. Advanced dashboard functions will display significant events related to the company such as news items about the company, about its competitors, about the industry and the economy. Other semantic-based events such as comments about the company in newsgroups, technical support bulletin boards, customer complaints, etc can be added to the schema. With the usage of advanced analytics and other technology

---

<sup>1</sup> Respectively, Ph. D. Student and William Von Minden Professor of AIS, faculty of Management, Rutgers University.

<sup>2</sup> Vasarhelyi, M. A., The Electronization of Business: a fundamental revolution. Working Paper, Faculty of Management, Rutgers University.

<sup>3</sup> <http://aicpa.org/assurance>

<sup>4</sup> CICA/AICPA. 1999. *Continuous Auditing*, Research Report. The Canadian Institute of Chartered Accountants, Toronto, Ontario.

<sup>5</sup> Bell, T. 7 Solomon Ira, *Auditing through a Strategic Lens*, KPMG, 1997.

<sup>6</sup> Kogan A., Sudit E.F. & Vasarhelyi M. A., "Continuous Online Auditing: A Program for Research", *Journal of Information Systems*, forthcoming Fall 1999.

interesting inferences may be drawn including “potential competitors” and "emerging issues<sup>7</sup>." which were automatically detected from the Internet using computer technology. Figures 1 and 2 illustrate the potential features of the media dashboard.

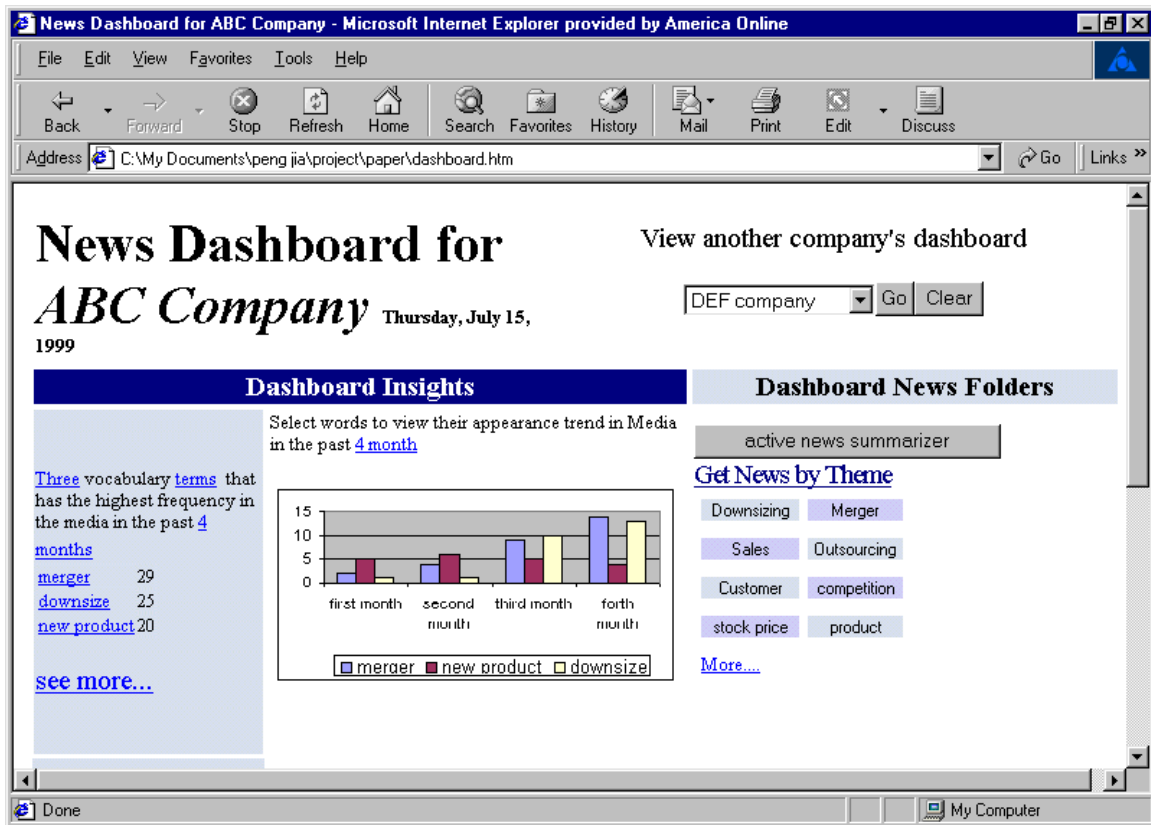


Figure 1 the first page of a potential media dashboard

<sup>7</sup> An emerging issue may be some item that is often mentioned by customers to company representatives or in e-mail received by the company, e.g. a competitive offer, sponsorship of a particular event, or a product defect.

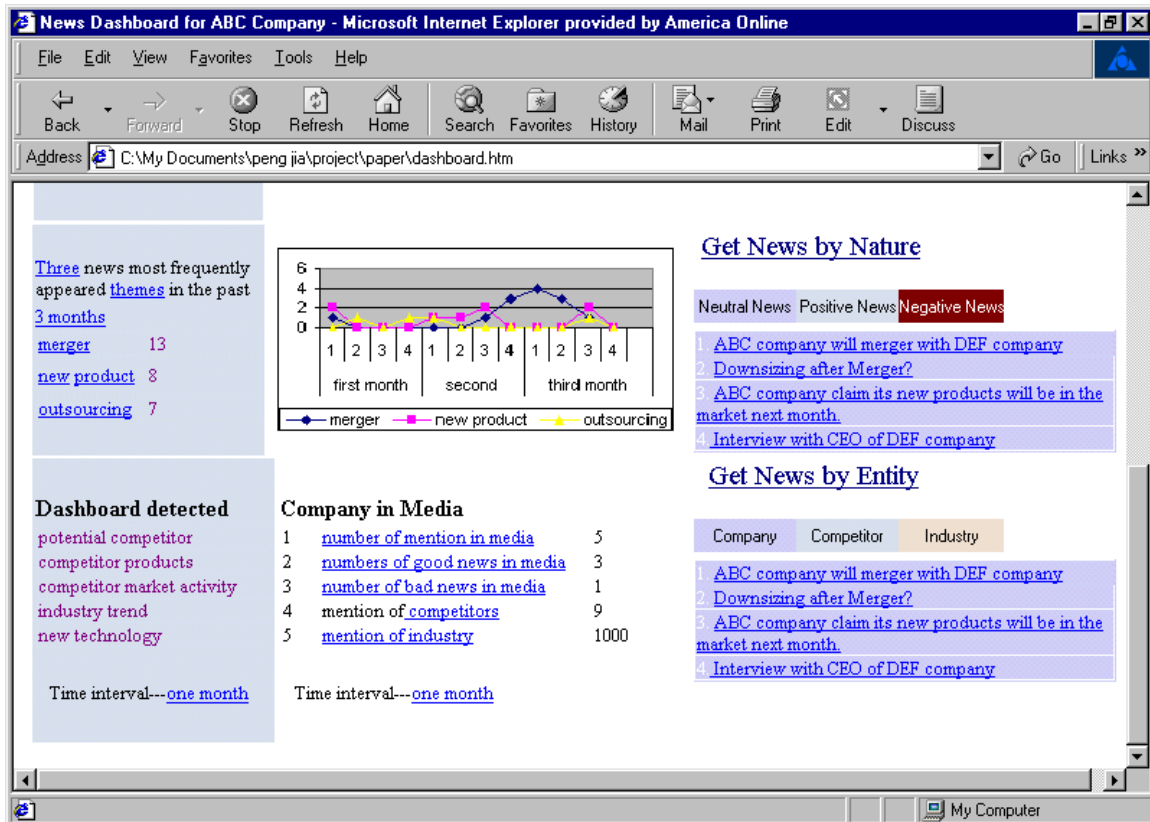


Figure 2 the second page of the media dashboard

Figures 1 and 2 illustrate the dashboard providing extensive but well organized media content to its users. The news provided in dashboard's news folder went through a filtering/selection process to ensure that they relate to a particular company (e.g. company ABC.) Furthermore, all the news items were classified by content, by nature (whether it is a good news or bad news), or by entity. This filtering was done by using computer text analysis techniques.

In "Get news by entity section", a company's competitors or industry's information can be displayed to facilitate audit comparison. Extra statistical information such as the most frequently mentioned terms in the news, within a particular time interval, provide a quick impression of the news' possible dominant content. Significant terms and news theme appearance trend charts illustrate the company's situation straightforwardly. All the news can be presented in a summary form, upon auditor specification.

### The potential value of on-line information to auditing services

The easy accessibility of World Wide Web and the proliferation of on-line news sources provide great opportunity for auditors to get large amounts of current information. On-line news sources are an unique channel for auditors to obtain valuable environmental information of entities and industries. Besides on-line news, Web sites, specially the company's competitor's Web site also provide information that can be used in many professional services such as strategic corporate consulting. The on-line textual information is simply a gold mine waiting to be exploited.

However, the amount of on-line information is often overwhelming to users. Using computer technology to automatically extract needed information and organize this information in a meaningful way, is a promising way to deal with this problem.

The next section discusses the systems architecture as well as a series of underlying text processing techniques.

## System architecture design and text processing techniques

### System architecture design

The system can be divided into 3 parts:

1. information collection — designed to collect news and other on-line information from Internet and filter out irrelevant information
2. information processing — automatic categorization of information, extraction of structured meta-data from unstructured texts and discovery of knowledge from meta-data using statistical methods.
3. result presentation — presenting text analysis results in user's preferred form.

The system architecture is shown in figure 3.

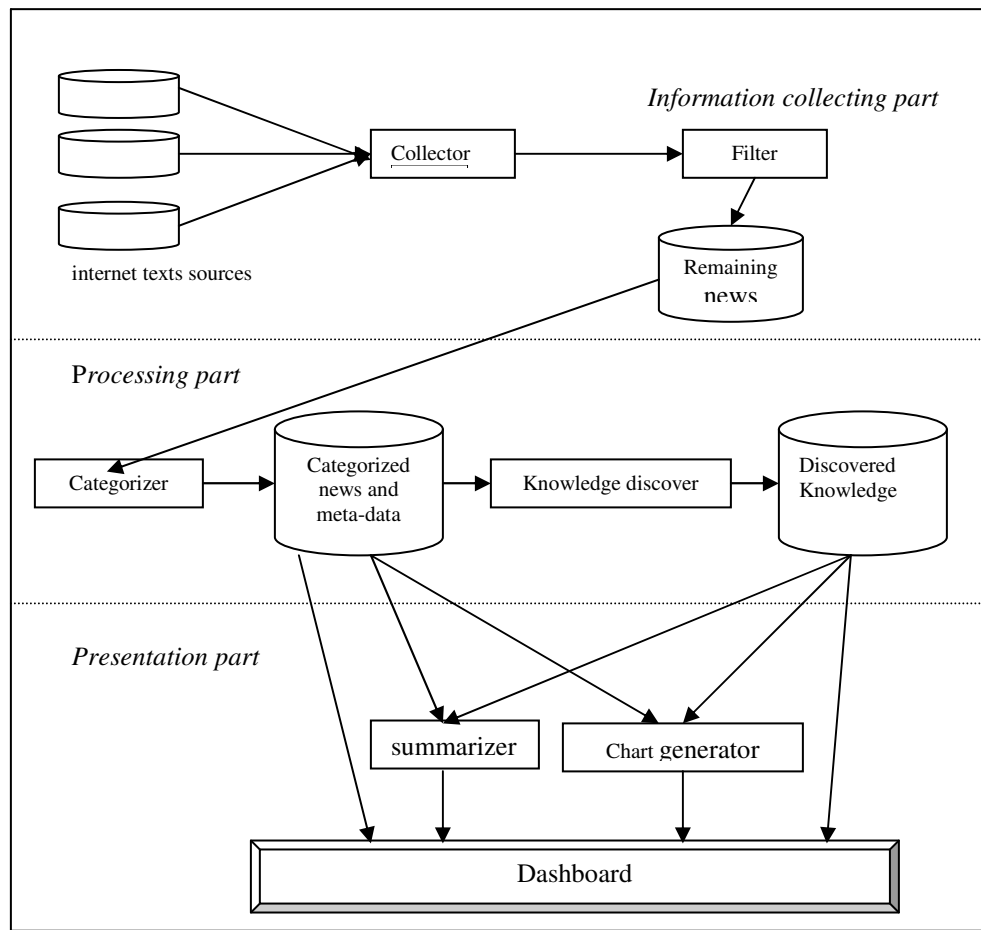


Figure 3. System architecture

### Text processing techniques

#### News collecting and filtering

The information **collector** knows how to communicate with on-line information sources and it gathers texts that include news and other Web site content from the Internet. For each company, the collector needs to gather, from auditor designated sources, news regarding the company itself, the company's competitors,

and the company’s industry. Extensive information sources, such as a competitor’s Web site, relevant third party sites, newsgroups, news streams need to be investigated and will be shown in the “Dashboard detected” section. The collector routinely gathers and accumulates information. Depending on whether the information source uses “pull” or “push” technology, the collector may actively grab or passively receive information from its source.

The information **filter** is employed to prevent irrelevant information from entering the system. Auditors need to manually construct a filtering profile specifying relevant news features. The profile can be a list of words that auditors want to appear or not appear in the texts. For example, if an auditor only want to keep news regarding a client and the client competitors, he/she can construct a profile with the client and competitors names in the preferred word list. Since it is a fact that almost all the news regarding a particular company mentions the company's name, it is reasonable to filter out an incoming news text whenever the text contains none of the names specified in the profile. More complex profiles could mix a client's name a key words for example; *LUCENT AND (MERGER OR ACQUISITION OR BUYOUT)*. More complex methods of profile building may entail tools where an auditor reads a number of news pieces (say 1000) and rates each piece as an 'include' or 'exclude.' Semantic analysis techniques, based on word frequency, distance, and preference are used then to build selective filtering.

**Information processing**

The key step of information processing is undertaken by the information categorizer, where structured information is extracted from unstructured news texts. Linguistic indexing, feature indexing, as well as keyword indexing can be used to extract information from news text. For example, news features such as business entity’s name mentioned in the text and semantically significant terms of the texts will be extracted and used, as meta-data that describing the text, to populate rows of a *categorized text database*. (Sample row of the database show in Figure 4). The categories a piece of news should belong to are then judged based on those meta-data. Accordingly, a list of category names will be annotated with the news and also will be stored into fields of the *categorized text database* (also see Figure 4).

Entity name	Entity Name	Term 1	Term 2	News Theme	Category 1	Category 2	News ID	
News	ABC	DEF	Merger	Down sizing	Merger	Merger	Positive news	Pointer to News Text

Figure 4. Sample row of the categorized text database

Texts can be categorized along several dimensions. Auditors may want to classify news by business entity, by theme (subject), by industry and/or by news’ nature (whether it is positive news or negative news). The larger the number of ways users want to categorize texts, the larger the number of fields in the *categorized text database*.

After the text’s meta-data were saved in the database, a series of statistical based algorithm will be employed to conclude or discover valuable information from the data. The final results, such as the most frequently used semantic terms of the text and its count will be saved in a database for future presentation. For example, to decide whether a piece of news positively or negatively reflects a company, the categorizer needs the following information:

1. A list of auditor predefined words, any of which, if appeared, indicates the positive nature of its context. Each word has a weight indicate its positive power.
2. A list of auditor predefined words, any of which, if appeared, indicates the negative nature of its context. Each word has a weight indicate its negative power.

According to this two lists, the categorizer looks in a piece of news for words that have positive or negative power and calculate these words' average weight to finally classify the news. Other forms of categorization base their operation on similar principle.

Many statistics present on the dashboard can be computed by summarizing meta-data in the database. Therefore, categorization is the main step that converts the un-measurable news into countable data.

### **Result presentation**

Some of the meta-data and knowledge from processing step can be displayed directly. Some need to be transformed by chart generator or text summarizer before show on the dashboard.

Automatic summarization of news will enable auditor to read less but get more. The summarizer first ranks sentences in a news text in descending order of relevance to the news topic. Then the most relevant sentences will be used to create the condensed version of news. Only when users indicate preference of looking only the news' summaries, the summarizer will be activated.

The chard generator uses data in the database to generate charts that visualize the data. It is enabled when the dashboard user wants more straightforward information.

### **System development**

Currently, a prototype dashboard system is under development at the Accounting Research Center of the Faculty of Management at Rutgers University. Around a Web tools for display and distribution IBM's *intelligent miner for text* provides part of the system components, the Perl language is used in constructing the information collector as well as integrating all components of the system together. Several different sources of data are being considered including pre-filtered news pieces being pushed by Pointcast and Yahoo, as well as drawing selection from a corpus of data from IP.

### **Conclusions**

Continuously monitoring business's operation and situation is an indispensable part of a successful continuous auditing system. The Internet has "exponentially increased the speed and availability" of information to auditors. One solution destined to be found to exploit enormous amount of on-line information. The media dashboard is a practical answer to this question. Its development could have valuable influence on the present auditing evolution towards a continuous auditing, by incorporating additional textual information into traditional information structures.