Agent Mediated Knowledge Management for Tracking Internet Behavior

Ira Haimowitz, Nuno Santo

Department of Business Technology/Information Science, Pfizer, Inc.
235 E. 42nd Street, New York, New York, 10017, USA
Haimoi@pfizer.com, Nuno.santo@pfizer.com

Abstract

We have developed a real-world business application performing agent-mediated knowledge management (AMKM) for the purposes of tracking Internet behaviors. Our application touches aspects of the following application areas: Analysis and design methods for AMKM, agents for group formation and awareness, agent-based workflow in the KM context, organizational implications of agent use in KM. We believe that this system illustrates several successful examples of agent-mediated KM in tracking and analyzing a complex set of drivers impacting Internet sites for the pharmaceutical industry and the application of Agents in the formalization of knowledge as key input for marketing business intelligence.

Business Challenge

Due to the dynamic nature of consumer Internet sites and the challenge of maintaining an active learning environment in which data is properly analyzed and fed back as formal business intelligence, a flexible knowledge management system is required. This becomes an increasingly complex problem as the number of sites to be tracked increases as well as the supporting marketing and technical organization for each site.

Our approach to these challenges involves centralizing data collection whereby various analytic processes are used to convert data into true marketing intelligence to be disseminated to the appropriate parties for decision making. For example, a marketing manager will require marketing intelligence that deals with the impact of a media campaign on the site whereas the technical team maintaining the site will require information on file server o security performance of the site. We later describe this contextual dissemination of data in a section on Agent Stereotypes.

Please note that our work never makes use of personally identifiable information. All analyses are done on an aggregate level, and no patterns are seen for individuals.

Defining the KM Problem

The central problem in building a KM system to support tracking Internet behavior lies in the development of key metrics by which to assess the performance of a visit. This is central to the KM system as each web site contains content and a site structure that is unique to every product thereby making each visitor behavior a distinct experience across sites of different products. There are several areas of performance that are vital for the supporting organization to best manage the web site. In the following sections, we outline our measurement ontology to deal with this data challenge followed by our KM ontology for classifying the discoveries based on these metrics.

Internet Behavior Measurement Ontology

We have defined an ontology of internet behaviors to track for our knowledge management system. Let us define a *visit* to a web site as a sequence of requested files $V = (f_1, f_2, \ldots, f_m)$. At the foundation of our behavior ontology are three types measurements about visits: segments, paths and errors.

1. A *segment* S_i is a set of pages $\{p_{i1}, \ldots, p_{in}\}$ belonging to a common content area, coupled with a threshold k_i . A web visit V belongs to segment S_i iff the file request sequence of V includes at least k_i pages from the set. For example, a visit could belong to "condition interest" segment if the visit sequence includes 2 of a set of 5 therapeutic condition-related pages.
2. A *path* is a meaningful predefined sequence of file requests $(g_1, g_2, \ldots, g_l)$. A visit V contains the path iff all of the files (g_i) occur in V, in the same order. An example path of 3 pages is (home page, brochure request form, form completed).
3. An *error* applies to visit V iff the visit terminates prematurely at some file f_i which cannot be served as the visitor requested. Examples are when the file f_i is not found (code 404), is not authorized to be seen (code 401), or the network connection times out (code 408).

We regularly track the percentage of visits on each web site that belong to segments or contain paths. We also track the absolute count of visits having errors. These percentages and counts are monitored using statistical process control (SPC) [Wheeler et. al 1992]. The time period in our SPC charts is most often daily for errors, and weekly for segments and paths. When a percentage is beyond the upper or lower control limits set by time-series forecasting, an alert is automatically sent to agents via e-mail.[1]

Agent Stereotypes

We encode our agents using *stereotypes* based on web site and job function. Thus the *marketing* stereotype agent for a given web site will receive e-mails about exceptional changes for segments and paths. The *technical* stereotype agent for that web site will receive emails about exceptional errors. Different agents may represent a person, if that person is cross-functional.

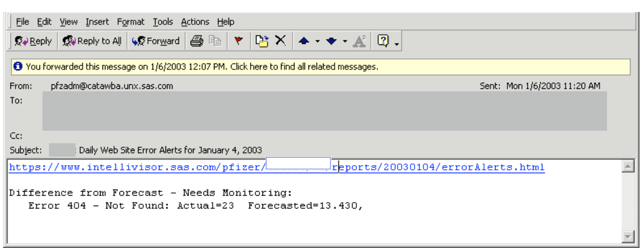
Tailoring communications using stereotypes is a decades old natural language technique, attributed to [Rich 1979]. It would be only a modest enhancement to allow people to build their own individualized agents picking and choosing what alerts they should be emailed. This would yield a more realistic mix [Sparck-Jones 1987] .

Tracking System Status and Example

Our automated tracking and e-mail alerting system has been implemented, serving over thirty users while tracking over ten product-related patient education as well as corporate Internet sites.

The web logs are analyzed daily using a leading commercial software tool. SAS IntelliVisor for Pharma [SAS 1991] is an Application Service Provider (ASP) solution that provides e-channel analysis to improve campaign strategies, marketing programs, and consumer understanding, which facilitates improvement in brand awareness activities.

Exceptional traffic-related alerts are sent daily to technical agents for their specific web sites. Exceptional business segment, pathing, and referral alerts are sent to marketing agents on a weekly basis.


File Edit View Insert Format Tools Actions Help

Reply | Reply to All | Forward

You forwarded this message on 1/6/2003 12:07 PM. Click here to find all related messages.

From: pfzadm@catawba.unx.sas.com Sent: Mon 1/6/2003 11:20 AM

To:

Cc:

Subject: Daily Web Site Error Alerts for January 4, 2003

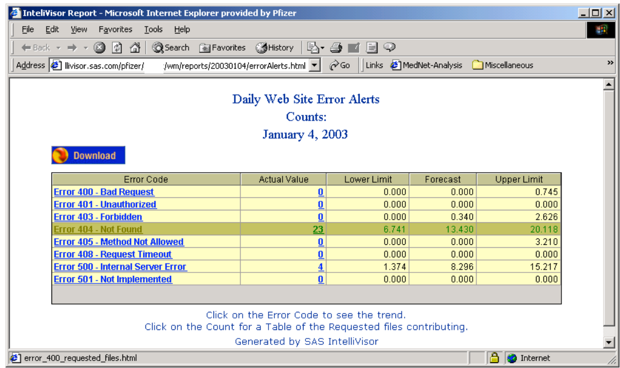
https://www.intellivisor.sas.com/pfizer/ reports/20030104/errorAlerts.html

```
Difference from Forecast - Needs Monitoring:
   Error 404 - Not Found: Actual=23  Forecasted=13.430,
```

[1] We also run similar alerts on more traditional *web traffic* measures like total visits, average visit time, etc.

Figure 1. Example e-mail alert send to "technical" agents for a given web site.

Above is an actual emailed alert to technical agents when a particular web site had an unusually high number of "Error 404 – File not Found" errors based on the time series forecast for the previous day. The email contains a hot-link to a web site hosted on SAS Intellivisor. A person receiving the email and clicking on the link goes to a technical agent alert summary page displayed in Figure 2.


InteliVisor Report - Microsoft Internet Explorer provided by Pfizer

Address llivisor.sas.com/pfizer/ /wm/reports/20030104/errorAlerts.html

Daily Web Site Error Alerts

Counts:

January 4, 2003

Download

Error Code	Actual Value	Lower Limit	Forecast	Upper Limit
Error 400 - Bad Request	0	0.000	0.000	0.745
Error 401 - Unauthorized	0	0.000	0.000	0.000
Error 403 - Forbidden	0	0.000	0.340	2.626
Error 404 - Not Found	23	6.741	13.430	20.118
Error 405 - Method Not Allowed	0	0.000	0.000	3.210
Error 408 - Request Timeout	0	0.000	0.000	0.000
Error 500 - Internal Server Error	4	1.374	8.296	15.217
Error 501 - Not Implemented	0	0.000	0.000	0.000

Click on the Error Code to see the trend.

Click on the Count for a Table of the Requested files contributing.

Generated by SAS IntelliVisor

error_400_requested_files.html

Figure 2. Technical agent alert summary page.

In this page a user can see an SPC profile for each technical metric on that alerted day. Listed are the metric, the actual value, the predicted value, and the lower and upper control limits. Clicking on a particular metric displays the process control chart illustrating the recent exceptionally high (or low) value. The chart appears below in Figure 3, demonstrating an unusually high recent value.

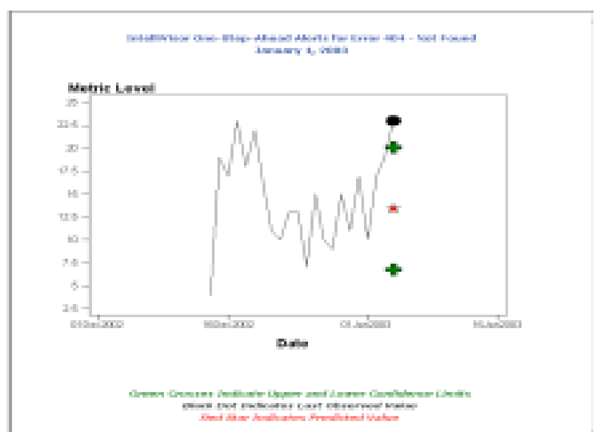


Figure 3. Statistical process control chart demonstrating latest days unexpectedly high "404 File not Found" errors.

A person reviewing this data decides whether or not to view other web log related analyses, to contact a colleague, or to take direct action. If the situation has ultimately led to a web site improvement or other business benefit, the user can submit an entry to the knowledge management system we have described earlier.

Organizational Interaction with KM System

In the current stage of our KM system as illustrated below in figure 4 and figure 5, we are manually inputting the discoveries of our agents that impact a decision related to managing the website both from a marketing and technical standpoint. In today's environment, we work closely with the marketing and technical teams and track which discoveries led to which decision in addition to projecting an economic benefit as a result of this new decision. To deal with the diversity of possible decisions and actions, we have developed a KM ontology that classifies decisions along several dimensions. They range from technical decisions such as improving the overall navigation of the site to which on-line campaign performed the best.

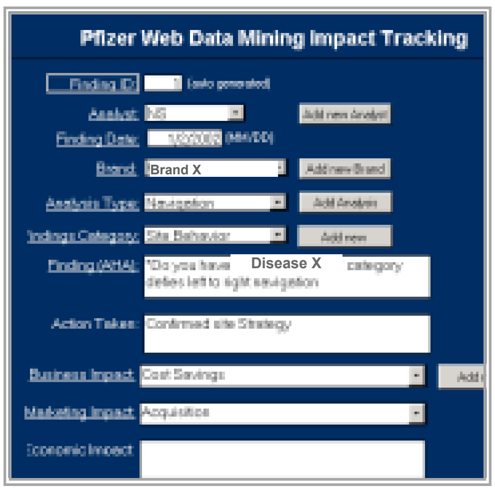


Figure 4. Example of our KM system data entry screen using our KM Ontology for inputting the impact of agent discoveries.

Business Impact

As we are actively tracking the decisions made as a result of our agent discoveries, we are able to collectively share these discoveries with our Internet management organization at large. We have developed a series of reports (see sample report in figure 5) that aggregate these discoveries and the decisions made across all the products sites as well as by the type of decision. These reports are shared with both the technical and marketing organization based on their stereotype. We have to date collected over 50 discoveries which have led to a range of significant decisions from an entire re-design of a web site, adding or removing content, to re-allocating the budget for various on-line media campaigns.

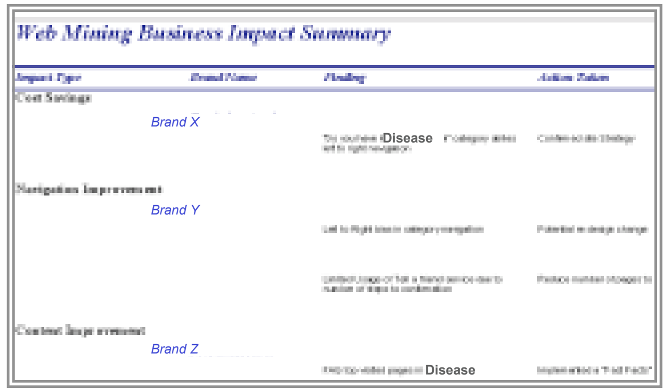


Figure 5. A sample report from our KM system listing the impact of agent findings.

Future Directions

Our agent-mediated knowledge management system has had great business impact, but could be made more automated to reduce the time delays and omissions inherent in of manual review of e-mailed alerts. We also would like agents to make automatic entries into the KM knowledge base. One route would be to make the agents more intelligent and program them to interact with each other, much like the buyer and seller agents of automated auctions as in [Jain et, al. 1999]. We can also reduce the number of alerts and yield a higher true-positive rate by using multivariate *trend templates* [Haimowitz and Kohame 1993].In addition, we plan to data mine the KM knowledge base to understand what web site attributes and alert types combine for biggest business benefit.

References

Haimowitz, I., Isaac, K. (1993), An Epistemology for Clinically Significant Trends

Jain, A.K., Aparicio, M., and Singh, M.P. (1999), Agents for Process Coherence in Virtual Enterprises, *Communications of the ACM*, vol. 42, no. 3, March 1999, pp. 62-69.

Rich, E. (1979), User Modeling via Stereotypes, *Cogitive Science*, vol 3, pp. 329-354.

SAS Institute Inc.(2001), SAS IntelliVisor for Pharma(SM), http://www.sas.com/solutions/intellivisor/pharma.html

Sparck-Jones, K. (1987) *Realism About User Modeling*, University of Cambridge Computer Laboratory Technical Report No. 111.

Wheeler, D. J., Chambers, D.S., Edwards, W. (1992), *Understanding Statistical Process Control*, SPC Press.