Analytics and Management of Collaborative Intranets

Peter Géczy, Noriaki Izumi, Shotaro Akaho, and Kôiti Hasida

National Institute of Advanced Industrial Science and Technology (AIST)
Tokyo and Tsukuba, Japan

Abstract. We present analytic framework for evidence-based management, design, and engineering of collaborative intranet environments. The analytics target elucidation of essential elements of human-system interactions. Temporal segmentation of human behavior in digital environments permits identification of crucial navigational points as well as higher order abstractions. Explorations of these elements provide fertile grounds for assessment of usability and behavioral characteristics that directly translate to actionable knowledge indispensable for improvements of collaboration portals. We extrapolate the analytic findings from a case study of a large scale collaborative organizational intranet; in order to identify three crucial domains facilitating alignment between observed evidence and best management and engineering practices.

Keywords: Collaborative Intranets, Web-based Portals, Analytics, Logs.

1 Introduction

"The overwhelming majority of organizations, however, have neither a finely honed analytical capability nor a detailed plan to develop one." [1] The absence of analytics deployment in organizations has limiting consequences on operational and management efficiency. Despite significant investments in organizational information and collaborative platforms, their usability remains low [2],[3].

The key to advancing usability of organizational collaborative information systems is the alignment of natural human behavior in electronic environments with the design and engineering. Investigation of human behavior in digital spaces has been attracting significant attention from corporate sector [4]. Corporations are eagerly collecting interactive and behavioral data about their web customers. Employed predictive modelling techniques of user behavior aim at converting visitors into customers—leading to increased revenues [5],[6]. Unfortunately, the usability improvements of their internal information and collaborative systems are undervalued.

We introduce an analytic framework addressing the needs of evidence-based management and design of collaborative intranets. The framework enables detailed analytics of human-system interactions. The analytics highlight knowledge pertinent to usability improvements and identification of high priority management domains.

E. Bertino and J.B.D. Joshi (Eds.): CollaborateCom 2008, LNICST 10, pp. 623–631, 2009.

2 Analytics Deployment

The implementation of analytics in organizational intranets is generally a multi-stage process. Four essential stages are: data collection, cleaning, preprocessing, and analysis. The processing flow is fundamentally sequential (as illustrated in Figure 1), however, it cyclically repeats whenever new data is collected. Ideally, on-the-fly analytics would be the most desirable. This requires availability of a computing power matching the data volumes and complexity of processing. In practice, periodical processing is often implemented (e.g. data collected during daytime is processed overnight).

Data collection in organizational intranets can be managed directly by the servers. Web servers have capability to record every arriving and served request to substantial detail. The information is stored in a predefined format—facilitating automated processing. The servers, however, record both human and machine generated requests. Hence the collected data requires cleaning.

Cleaning should eliminate the machine generated traffic (e.g. automated network monitors checking responsiveness of serves, crawlers, indexers, etc.) and preserve the human generated traffic. Machine traffic is often voluminous. Cleaning may substantially reduce the data volume—easing the following processing.

Preprocessing prepares clean data for analytic processing. Server log records contain several information fields. The log records need to be parsed for individual fields. The extracted information should be suitably organized; preferably structured and databased for further analysis.

Analysis targets various explorations of human behavior in intranet environments; such as usability analysis, user profiling, collaborative filtering, etc. The analytics play a vital role in intranet management, personalization, service deployment, and business process (re-)engineering.

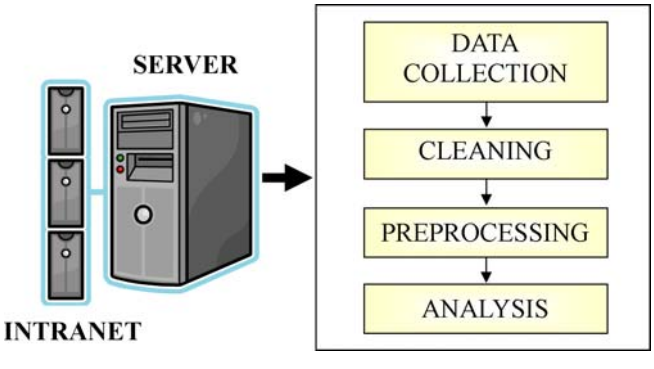

Fig. 1. Depiction of sequential intranet analytics implementation consisting of four stages: data collection, cleaning, preprocessing, and analysis

3 Analytic Framework

Human dynamics in electronic environments have temporally specific features [7]. Users exhibit periods of activity followed by longer inactivity periods. The human-web interactions can be segmented with respect to these temporal indicators. The temporal segmentation framework of human browsing behavior has been introduced in [2]. We concisely recall relevant constructs.

The page transition sequences are segmented into *sessions* and *subsequences*. The segmentation is illustratively outlined in Figure 2. The sessions represent more complex tasks undertaken by the users. They are further divided into the subtasks indicated by the subsequences. Consider the sequence of the form: $\{(p_i, d_i)\}_i$ where p_i denotes the visited page URL_i and d_i denotes a delay between the consecutive views $p_i \rightarrow p_{i+1}$. **Session** is a sequence $B = \{(p_i, d_i)\}_i$ where each $d_i \leq T_B$. **Subsequence** $S = \{(p_i, d_i)\}_i$ consists of pairs (p_i, d_i) with delays satisfying the conditions: $d_i \leq T_S$.

Pertinent issue in segmenting the human browsing interactions into sessions and subsequences is the appropriate determination of the separating delays—the values of T_B and T_S. Explorations of student web behavior revealed that their browsing sessions last on average 25.5 minutes [8]. Analysis of knowledge worker browsing interactions on the organizational intranet portal exposed longer session duration: 48.5 minutes on average [2]. The study used empirically determined minimum inter-session delay $T_B = 1\ hour$. The subsequence delay separator T_S was calculated dynamically as an average delay in the session bounded from below by 30 seconds.

Consider the following example of interactions on the organizational intranet. At the beginning, a user logins into the intranet system—subsequence 1. Successful login process will lead to a display of the initial portal page. Navigating from the opening page, a user accesses a bulletin board of an organizational interest group he/she belongs to—subsequence 2. On the bulletin board a user locates and reads the latest announcement concerning a group meeting—subsequence 3. Then he/she leaves the environment for another work-related activity.

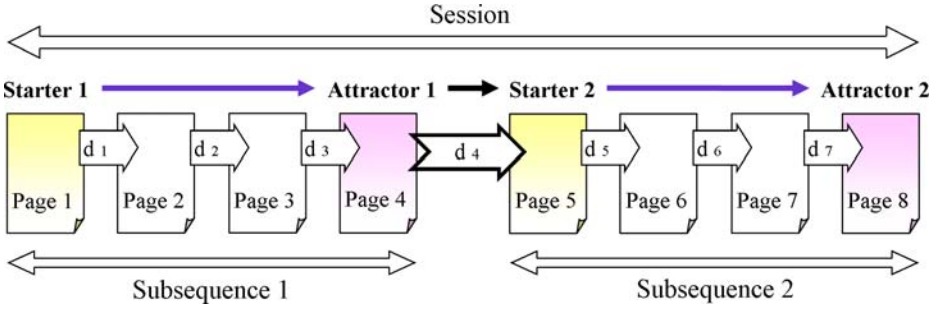

Fig. 2. Segmentation of web interactions and identification of navigation points

The example of user's interactions on an organizational intranet portal illustrates a potentially typical session consisting of three subsequences. Each subsequence corresponds to a distinct task: 1–login, 2–accessing a bulletin board, 3–locating and reading an announcement. Each subsequence exhibits relatively dynamic transitions from the initial page to the target. At the target page the user's attention is required. The attention takes time. Hence, there are longer delays recorded prior to initiating the next browsing task.

Important navigation points are the points where users initiate their browsing actions—*starters*, the resources they target—*attractors*, and the single action behaviors—*singletons*. **Starters** are the first points of subsequences or sessions with length greater than 1. **Attractors** are the last points of subsequences or sessions with length greater than 1. **Singletons** are the points of subsequences or sessions with length equal to 1.

In the formerly presented example of user interactions on the intranet portal, the *starter* of the first subsequence is the login page, and the *attractor* is the opening portal page (after successful login). Singletons outline a single action behaviors surrounded by longer delays. They generally relate to the use of hotlists, such as history or bookmarks, where users access the desired resource directly without proceeding throughout several transitions [9].

Subsequences constitute the elemental tasks and browsing segments. Users can take various navigational paths from the initial point to the target. Since the intermediate points between the starter and the attractor are only transitional, tracking the multitudes of all navigational pathways may be a waste of resources. It is more efficient to focus on the abstractions of the segments. Starter-attractor pairs, denoted as **SA elements**, are suitable abstractions of subsequences. The connecting elements of subsequences are represented by attractor-starter pairs. Formation of more complex browsing patterns can be observed from sequences of SA elements and connectors.

Segmented user behavior, into sessions and subsequences, identification of important navigation points, and higher order abstractions provide valuable information enabling analysis of collaborative intranet environments. Analytic findings often translate into actionable knowledge for managers, administrators, and designers. This enables continual improvement of the existing services and identification of the novel ones.

Various analytics, measures, and metrics can be derived based on the presented framework. In practice, it is often beneficial to explore and analyze the usability of digital environments. Exploratory analysis of web log data, based on the introduced framework, exposes valuable findings related to how users utilize the environment, which resources users find useful, which resources and processes pose difficulties in access or execution, etc. This study explores the essential statistical analytics.

4 Case Study

The investigated organizational collaborative intranet is a large-scale system implemented at The National Institute of Advanced Industrial Science and

Technology. The system is significantly complex and distributed. The web core comprises of six servers connected to the high-speed backbone in a load balanced configuration. Users access the system via local subnets with infrastructures ranging from high-speed optical to wireless.

The collaborative intranet portal incorporates a large number of resources and services in a decentralize and distributed manner. A rich spectrum of resources span across documents in various formats, multimedia, software, etc. The extensive range of implemented services support the organizational business processes; cooperation with industry, academia, and other institutes; internal collaboration; localization of resources, etc. Local networking and blogging services within the organization are also implemented. Visible web space is in the excess of 1 GB, and deep web space is considerably larger, however, it is difficult to estimate its size due to the distributed architecture and constantly changing back-end data.

The collaboration intranet portal manages considerable traffic. The peak traffic is primarily during the working hours. There are also few cyclical traffic peaks. Collected web log data by the servers is voluminous and contains details of portal utilization and users' browsing features. The web logs contain records of both human and machine generated traffic. Cleaning, preprocessing, and segmentation of human-web interactions has been presented in [2]. The resulting working data, together with the essential intranet statistics, are described in Table 1.

Table 1. Information and basic data statistics of the organizational intranet portal

Web Log Volume	∼60 GB
Average Daily Volume	∼54 MB
Number of Servers	6
Time Period	1 year
Log Records	315 005 952
Resources	3 015 848
Sessions	3 454 243
Unique Sessions	2 704 067
Subsequences	7 335 577
Unique Subsequences	3 547 170
Valid Subsequences	3 156 310
Unique Valid Subsequences	1 644 848
Services	855

5 Analytic Findings and Managerial Implications

Exploratory analysis of data revealed significant statistical usability pattern. Essentially all analyzed aspects of user interactions on the large-scale organizational collaborative intranet indicated long tailed characteristics. The typically observed pattern is depicted in Figure 3. It displays the access statistics of portal services.

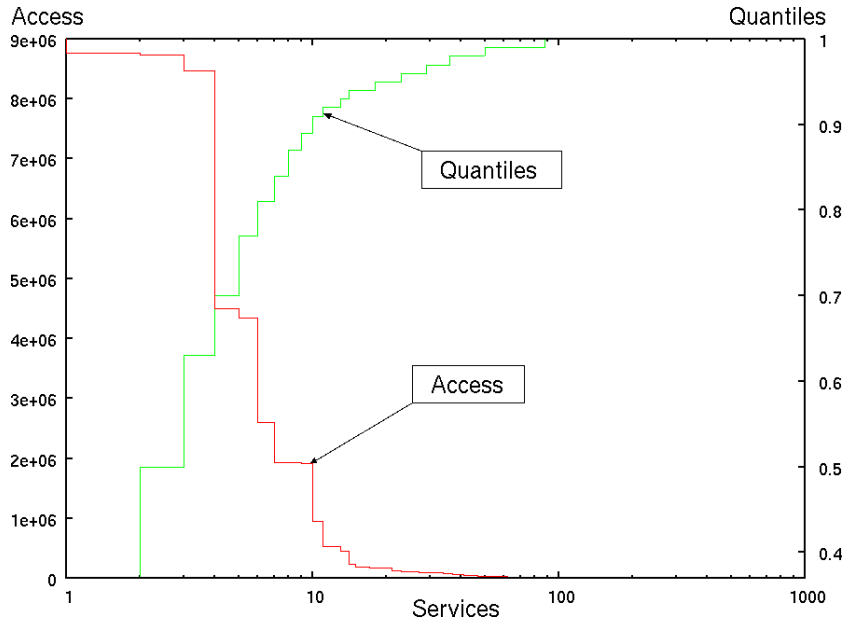

Fig. 3. Histogram and quantile characteristics of services access. X-axis is in logarithmic scale. Left y-axis refers to access histogram and right y-axis refers to quantiles.

It is noticeable that relatively few services were frequently used (note that x-axis is in a logarithmic scale, in order to visualize the head of the distribution). Only approximately ten services (out of 855) were frequently used. Over half of the services were accessed less than ten times. Top three services, that is 0.35%, accounted for 50% of use. These were: bulletin boards (containing organizational and other announcements), attendance service (recording, verifying, and altering the presence, business trip, and holiday records), and information service from human resources division (finding information about the members of the organization, such as e-mail, phone numbers, location, etc.). Bulletin boards were accessed most frequently. It indicates that users were generally interested in organizational and personal announcements.

Frequently used services should be easily accessible, highly optimized and suitably personalized. Unfortunately, majority of organizations do not pay sufficient attention to analytics, optimization, and personalization of internal web-based environments that should foster collaboration and higher working efficiency. They often use one-fit-all templates resulting in low portal usability. Following are concise highlights of further principal findings.

– **Underutilization of resources and services.** Evidence of service underutilization has been described in the former paragraphs. Analogous situation has been observed also in the case of other resources. Knowledge workers have generally utilized a small set of starters and attractors. The size of the

starter and attractor sets amounted for approximately 3.84% and 9.55% of the total navigation points, respectively.

- **Few useful resources.** Number of unique single user actions—represented by singletons–have been minuscule. The unique singletons have accounted for only 1.92% of navigation points. Only about ten navigation points have been found substantially useful by users to be included in their hotlists and recurrently accessed.
- **Short attention span.** The average attention span at the target has been observed to be only approximately 6.5 minutes. This has been indicated by the detected delays between subsequences.
- **Rapid transitions.** The peak interval of subsequence duration has been between four to six seconds. During this time, the users have made four to five page transitions–on average. This implies rapid pace—approximately one second per page transition.
- **Task subdivision.** The observed peak interval of number of subsequences in session has been $< 2, 4 >$. The browsing sessions have contained, on average, three subsequences. This suggests that the users have divided their more complex browsing and interaction tasks into three simpler and less complex subtasks.
- **Pattern formation.** Knowledge workers have formed elemental and complex interaction patterns. This has been indicated by a relatively small number of frequently repetitive SA elements and connectors. Frequent repetition of elemental patterns interconnected by frequent connectors exposes formation of more complex patterns.
- **Common browsing pattern.** Smaller number of starters have repeated substantially more frequently than the adequate number of attractors. The starter-attractor frequency ratio has been approximately one-to-ten. This implies that the general browsing strategy reflected the knowledge of the starting navigation point and familiarity with the traversal pathway to the target.
- **Behavioral habituation.** Knowledge worker browsing behavior has gradually habituated. At early stages of use, the resource access has been more 'ballanced'. Users' familiarity with the environment has resulted in gradually shrinking set of frequently accessed resources. Furthermore, they have exhibited faster navigational transitions.

The management of the organizational collaborative intranet portals should inevitably take into account the presented analytic evidence. It exposes the actionable knowledge related to several crucial usability aspects. The implications of the findings extend to various managerial domains. The three essential domains to be addressed are highlighted in the following paragraphs.

Process Engineering. Organizations are increasingly transferring internal collaborative and business processes to their information platforms. Design and engineering of business and collaborative processes should be aligned with the fundamental elements of human interactions. The elements of human interactions are rarely accounted for during the design and engineering. This decreases the

user working efficiency and induces the negative experience effects. It is often the case that the actual implementations of processes result in multistage extended human-system interactions. The analytic findings indicate that the users naturally divide their browsing tasks into three subtasks–on average. This implies that the extensive and prolonged processes should be re-engineered, so that they can be segmented into approximately three sub-processes. Furthermore, each stage should require less than seven minutes of human attention.

Assistance Services. Proficiency in use of organizational collaborative platforms requires practice–which can be time and resource consuming. To accelerate the learning, the users should be systematically assisted during their interactive experiences. Conversely, the extensive use and increased proficiency is linked with habituation and lack of exploratory behavior. Suitable assistance services can facilitate steeper learning curves and improved attention management. Assistance services should at least incorporate technologies such as collaborative filtering, recommendations, and localized search. The effective search enables fast location of closely relevant resources. The recommendation systems utilize the knowledge of use histories, preferences, and other analytics to offer suitable recommendations on resources and services of potential interest to the users [10]. The collaborative filtering technologies leverage the correlations among 'like-minded' users in providing assistance and recommendations [11],[12]. All these services can efficiently (re-)use the analytics.

Personalization. Different users have different interaction and browsing styles. Conventionally, the interactivity of organizational information systems with the human users has followed the one-fit-all style; possibly with exceptions of few optional adjustments. This interfacing is inadequate and often leads to underutilization. Human interactivity should be approached in a personalized manner. Personalization should be realized essentially on two levels: group and individual. The group personalization underlines the interface adjustments for a set of users with similar behavioral and/or working characteristics. The personalization on the individual level allows delivery of personalized interfaces for individual users. Personalization is closely linked with profiling. Each user, or group, should have a well formed profile incorporating the essential elements of human-system interactions. Based on the profile characteristics, the interface adjustments should be carried out adaptively and on-the-fly.

6 Conclusions

We presented the significant analytic findings of a large scale organizational collaborative intranet case study. The analysis revealed important usability and behavioral characteristics of users that translate into actionable knowledge for managers, administrators, and designers of collaborative intranet systems. The findings advocate effective deployment of appropriate system analytics as a necessary precursor for evidence-based management. The analytics provide a vital source of knowledge for decision making. Efficient collaborative intranet platforms require alignment of design, engineering, and management of the systems with

usability and behavioral characteristics of users. The alignment demands covering the three high priority domains: process engineering, assistance services, and personalization. Further extensions and improvements of collaborative intranet environments can suitably utilize and build upon these three essential bases.

References

1. Davenport, T.H., Harris, J.G.: Competing on Analytics: The New Science of Winning. Harvard Business School Press, Boston (2007)
2. Géczy, P., Akaho, S., Izumi, N., Hasida, K.: Knowledge worker intranet behaviour and usability. Int. J. Business Intelligence and Data Mining 2, 447–470 (2007)
3. Huntington, P., Nicholas, D., Jamali, H.R.: Website usage metrics: A re-assessment of session data. International Journal of Information Processing and Management 44(1), 358–372 (2008)
4. Petre, M., Minocha, S., Roberts, D.: Usability beyond the website: an empirically-grounded e-commerce evaluation for the total customer experience. Behaviour and Information Technology 25, 189–203 (2006)
5. Park, Y.-H., Fader, P.S.: Modeling browsing behavior at multiple websites. Marketing Science 23, 280–303 (2004)
6. Moe, W.W.: Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. Journal of Consumer Psychology 13, 29–39 (2003)
7. Barabasi, A.-L.: The origin of bursts and heavy tails in human dynamics. Nature 435, 207–211 (2005)
8. Catledge, L., Pitkow, J.: Characterizing browsing strategies in the world wide web. Computer Networks and ISDN Systems 27, 1065–1073 (1995)
9. Thakor, M.V., Borsuk, W., Kalamas, M.: Hotlists and web browsing behavior–an empirical investigation. Journal of Business Research 57, 776–786 (2004)
10. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17, 734–749 (2005)
11. Symeonidis, P., Nanopoulos, A., Papadopoulos, A.N., Manolopoulos, Y.: Collaborative recommender systems: Combining effectiveness and efficiency. Expert Systems with Applications 34(4), 2995–3013 (2008)
12. Jin, R., Si, L., Zhai, C.: A study of mixture models for collaborative filtering. Information Retrieval 9, 357–382 (2006)