Mining the Web for Access Decision-Making in Secure Spaces

Shun Hattori
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo, Kyoto 606–8501, Japan
Email: hattori@dl.kuis.kyoto-u.ac.jp

Tanaka Katsumi
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo, Kyoto 606–8501, Japan
Email: tanaka@dl.kuis.kyoto-u.ac.jp

***Abstract*—In public spaces, there are a number of different contents such as visitors, physical information resources, and virtual information resources via their embedded output devices (e.g., displays and speakers). Therefore, we might unexpectedly enter the public spaces that have our unfavorable characteristics (e.g., dismal and dangerous) and/or our unwanted information. For this problem, we have proposed a model and architecture of "Secure Spaces", which provide access control over public output devices and entry control over electrical lockable doors to prevent visitors from entering the public spaces that have their unfavorable characteristics and/or their unwanted information according to their access policies for spaces or information, that is, what spaces or information they do not want to access. This paper tackles how to extract information for making access or entry decisions in Secure Spaces from very large text corpora such as the Web to enables users to more flexibly specify their access policies by keyword-based expressions in practice.**

I. Introduction

In recent years, how to make physical spaces smarter has become one of the hottest topics in the research field of ubiquitous/pervasive computing. Smart Spaces [6] are often physically isolated environments such as rooms, which are made smart by various information communication technologies. They would be much more convenient for information access in the future. Meanwhile, information security has also become very significant in any situation, especially in public places such as indoor work places, educational facilities, healthcare centers and so on. The amount of physical or virtual information resources which should be protected in the physical world grows exponentially.

Physical environments are becoming smart but not always secure. When a virtual (computational) information resource is requested to access by a user via an output device, conventional access control systems make a decision on whether the user should be granted or denied to access the resource based on its access policies and surely enforce the access decision. However, even if the requester is authorized by it, it should not be immediately offered to her via the output device, because there might be its unauthorized users as well as the authorized requester around the output device, especially in public places. Meanwhile, when a user enters a physical environment, the user might hate its real characteristics (e.g., degrees of dismal and danger) and/or be forced to access her unwanted information resources unexpectedly.

In our previous works [1], [2], we have aimed at making Smart Spaces always secure in the real world, and defined Secure Spaces as physical isolated environments where any resource is always protected from its unauthorized objects with respect to information security, that is, any information resource is always protected from being accessed by its unauthorized visitors, while any visitor is always protected from being pushed her unwanted information resources on.

In this paper, we propose a method to extract information for making access or entry decisions in Secure Spaces from very large text corpora such as the Web, especially the Weblog, and also improve our previous architecture of Secure Spaces and mechanism of space entry control by adding the concept of the Weblog sensor, in order to enables users to specify their access policies by keyword-based expressions.

The remainder of this paper is organized as follows. Section 2 improves our previous architecture of Secure Spaces and mechanism of space entry control. Section 3 proposes a method to mine the Weblog for access decision-making in Secure Spaces. Section 4 shows several experimental results to validate the method. Section 5 introduces related works. Finally, we conclude this paper in Section 6.

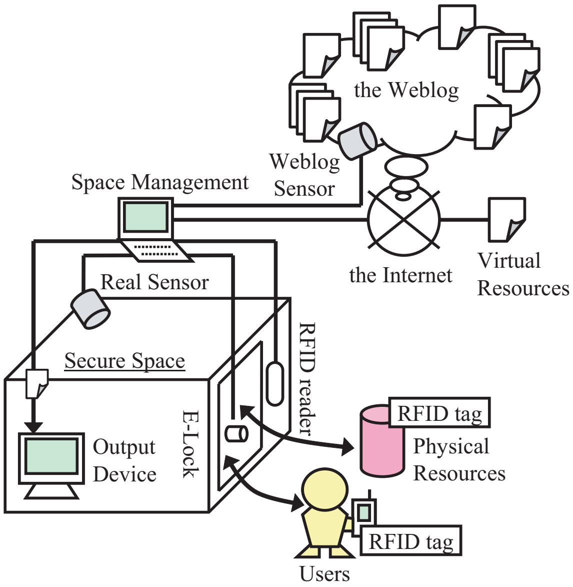

Fig. 1. Architecture of Secure Spaces

II. Secure Spaces

In this section, we improve our previous architecture of Secure Spaces and mechanism of space entry control in order to enable users to more flexibly specify their access policies for not only information resources but also Secure Spaces.

A. Architecture

To build Secure Spaces in the real world by using our space entry control based on their dynamically changing contents such as their visitors, physical information resources and virtual information resources via their embedded output devices, each Secure Space requires the following facilities (shown in Figure 1 in the previous page).

- **Space Management**: is responsible for managing a Secure Space, that is, for constantly figuring out its contents such as its visitors, its embedded physical resources and virtual resources outputted via its embedded output devices and also for ad-hoc making an authorization decision on whether an entry request to enter it by a visitor or a physical/virtual resource should be granted or denied.
- **User/Object Authentication**: is responsible for authenticating what physical entity such as a user or a physical resource requests to enter or exit the Secure Space (e.g., by using Radio Frequency IDentification or biometrics technologies) and also for notifying it to the space management.
- **Electrically Lockable Door**: is responsible for electrically locking or unlocking itself, that is, for assuredly enforcing entry control over physical entities such as users and physical resources, according to instructions by the space management.
- **Physically Isolating Opaque Wall**: is responsible for physically isolating inside a Secure Space from outside there with regard to information access, that is, for validating our assumption that any user inside a Secure Space can view any resource inside the Secure Space while any user outside the Secure Space can never any resource inside the Secure Space.

In addition to the above four facilities that our previous system architecture also has,

- **Real Sensor**: is responsible for physically sensing inside a Secure Space for its real characteristics to make access decisions in the Secure Space and also for notifying the sensor data stream to the space management. For example, thermometers, hygrometers, (security) cameras, and so forth. Note that a Secure Space does not always equip all of the real-sensing devices necessary to check whether or not the Secure Space meets a user's preference.
- **Weblog Sensor**: is responsible for logically sensing the Weblog for the approximate characteristics of each Secure Space to make access decisions in the Secure Space and also for notifying the Web-mined data to the space management. Note that any Secure Space does not have to equip the extra devices.

B. Space Entry Control

When a user requests to enter a Secure Space, our entry control system will make an entry decision on whether the entry request should be granted or denied, by checking whether or not the requester is granted to access by all information resources inside the Secure Space and whether or not the Secure Space itself as well as all information resources inside the Secure Space are granted to be accessed by the requester, in order to protect her preference as well as information security for all contents of the Secure Space.

If the Secure Space is not granted to be accessed by the requester because its real characteristics (e.g., degrees of dismal or danger) which the requester has not yet experienced by herself are unfavorable for the requester, our entry control system has only one approach of preventing the requester from entering the Secure Space.

If the requester is not granted to access by at least one physical information resource inside the Secure Space or if at least one physical information resource inside the Secure Space is not granted to be access by the requester, our entry control system has also only one approach of preventing the requester from entering the Secure Space (shown in Figure 2).

Meanwhile, if the requester is not granted to access by at least one virtual information resource via an output device embedded inside the Secure Space or if at least one virtual information resource is not granted to be access by the requester, our entry control system has two approaches of not only preventing the requester from entering the Secure Space but also permitting the requester to enter the Secure Space after revoking the virtual information resource (shown in Figure 3).

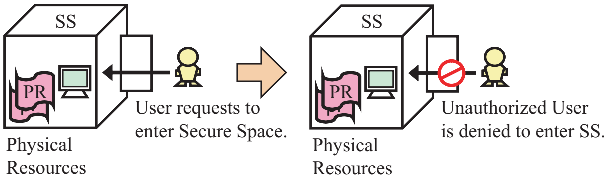
SS
PR
User requests to enter Secure Space.
Physical Resources
SS
PR
Unauthorized User is denied to enter SS.
Physical Resources

Fig. 2. Space entry control over user for physical resources

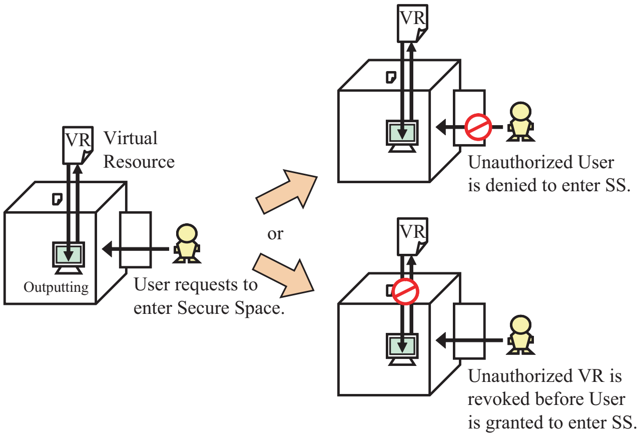
VR
Virtual Resource
Outputting
User requests to enter Secure Space.
or
VR
Unauthorized User is denied to enter SS.
VR
Unauthorized VR is revoked before User is granted to enter SS.

Fig. 3. Space entry control over user for virtual resources

Our formalized model of space entry control for Secure Spaces is summarized as follows.

- **Secure Space**, **Users**, and **Physical/Virtual Resources**: are atomic model entities and are denoted by S, U, PR, and VR respectively.
- Our model functions to keep up on the set of ad-hoc entities in each Secure Space and evaluate the weight of its dynamically changing contents:
  - $\mathtt{cu}$: $\mathsf{S} \rightarrow 2^{\mathsf{U}}$, is a function mapping each Secure Space s, to the set of its Containing Users $\mathtt{cu}(s)$.
  - $\mathtt{cpr}$: $\mathsf{S} \rightarrow 2^{\mathsf{PR}}$, is a function mapping each Secure Space s, to the set of its Containing Physical Resources $\mathtt{cpr}(s)$.
  - $\mathtt{cvr}$: $\mathsf{S} \rightarrow 2^{\mathsf{VR}}$, is a function mapping each Secure Space s, to the set of its Containing Virtual Resources $\mathtt{cvr}(s)$.
  - $\mathtt{w}$: $\mathsf{S} \times 2^{\mathsf{U}} \times 2^{\mathsf{PR}} \times 2^{\mathsf{VR}} \rightarrow \mathcal{R}$, is a function mapping a set of contents $\mathtt{cu}(s)$, $\mathtt{cpr}(s)$ and $\mathtt{cvr}(s)$ in each Secure Space s, to its evaluated Weight $\mathtt{w}(s, \mathtt{cu}(s), \mathtt{cpr}(s), \mathtt{cvr}(s))$.
- **User's Access Policy**: is an access policy defined as a 2-tuple of a User and her qualified Secure Space or Physical/Virtual Resource.

$$\mathsf{UAP} \subseteq \mathsf{U} \times \mathsf{O}, \text{ where } \mathsf{O} = \mathsf{S} \cup \mathsf{PR} \cup \mathsf{VR}.$$

- **Physical/Virtual Resource's Access Policy**: is an access policy defined as a 2-tuple of a Physical/Virtua Resource and its qualified User.

$$\mathsf{PRAP} \subseteq \mathsf{PR} \times \mathsf{U}, \text{ and } \mathsf{VRAP} \subseteq \mathsf{VR} \times \mathsf{U}.$$

- **Authorization for Users**: grants an entry request that a User u requests to enter a Secure Space s, if and only if any of the Secure Space and contents inside there grants the user to access itself and is granted to be accessed by the user or if the weight $\mathtt{w}(s, u)$ in case of granting the user to enter there after revoking any virtual resource inside there which denies the user to access itself or is denied to be accessed by the user is higher than the weight $\mathtt{w}(s)$ in case of denying the user to enter there.

$$\begin{aligned}&\forall u \in \mathsf{U}, \forall s \in \mathsf{S}, \mathtt{authU}(u, s) = \mathtt{grant}\\ &\Leftrightarrow ((u, s) \in \mathsf{UAP} \wedge \mathtt{apr}(s, u) = \mathtt{cpr}(s))\\ &\quad \wedge \{(\mathtt{avr}(s, u) = \mathtt{cvr}(s)) \vee (\mathtt{w}(s, u) \geq \mathtt{w}(s))\},\end{aligned}$$

where $\mathtt{au}(s, u)$, $\mathtt{apr}(s, u)$ or $\mathtt{avr}(s, u)$ is the Assumptive set of Users, Physical Resources or Virtual Resources inside the Secure Space s after granting the User u to enter there and regulating its contents to keep secure.

$$\begin{aligned}\mathtt{au}(s, u) &= \mathtt{cu}(s) \cup \{u\},\\ \mathtt{apr}(s, u) &= \{pr \in \mathtt{cpr}(s) | (pr, u) \in \mathsf{PRAP}\\ &\qquad \text{and } (u, pr) \in \mathsf{UAP}\},\\ \mathtt{avr}(s, u) &= \{vr \in \mathtt{cvr}(s) | (vr, u) \in \mathsf{VRAP}\\ &\qquad \text{and } (u, vr) \in \mathsf{UAP}\},\\ \mathtt{w}(s) &= \mathtt{w}(s, \mathtt{cu}(s), \mathtt{cpr}(s), \mathtt{cvr}(s))\\ \mathtt{w}(s, u) &= \mathtt{w}(s, \mathtt{au}(s, u), \mathtt{apr}(s, u), \mathtt{avr}(s, u)).\end{aligned}$$

III. Weblog Sensor

The Weblog sensor mines Weblog documents for access decision-making in each Secure Space, that is, for its approximate characteristics necessary to check whether or not the Secure Space meets a user's preference.

Let's suppose that a user specifies her access policies for Secure Spaces by using a keyword-based expression kw. For example, the user does not want to enter any Secure Space whose degree of "Yuutsu (dismal)" or "Kiken (danger)" is enough high. And that the user requests to enter a Secure Space s at a current time t_2. In order to extract its approximate degree of a keyword-based expression kw for the Secure Space s from Weblog documents by using text mining techniques, we have to convert its logical identity s to some linguistic description, that is, its place-name sn. Note that any Secure Space can be assigned multiple place-names to. For example, the Secure Space shown in Figure 4 is annotated by "Gion" and "Kyoto" (which geographically contains "Gion").

In general, an approximate degree of a Secure Space s for a keyword-based expression kw at a time t_2 is defined as

$$\begin{aligned}&\text{WeblogSensor}_{t_2}(s, kw) :=\\ &\sum_{sn \in N(s)} \sum_{t_1 \in T(t_2)} \text{weight}_{[t_1, t_2]}(sn, kw) \cdot \frac{1}{\text{area}(sn)} \cdot \frac{1}{t_2 - t_1},\end{aligned}$$

where $N(s)$ stands for a set of place-names assigned to a Secure Space s, $T(t_2)$ stands for a set of time points before a current time t_2 (e.g., a time point t_1 is defined as a day/month before t_2), $area(sn)$ stands for the geographical area of a place-name sn, and

$$\begin{aligned}\text{weight}_{[t1, t2]}(sn, kw) &:= \frac{\Pr_{[t1, t2]}(sn|kw)}{\Pr(kw)},\\ \Pr_{[t1, t2]}(kw|sn) &:= \frac{\text{bf}_{[t1, t2]}(sn \wedge kw)}{\text{bf}_{[t1, t2]}(sn)},\\ \Pr(kw) &:= \frac{\text{bf}(kw)}{N},\end{aligned}$$

where $\text{bf}_{[t1, t2]}(q)$ stands for the number of searched Weblog documents by submitting a query q with an optional time-interval $[t1, t2]$ to such a Weblog search engine as Yahoo! Blog Search [3] and N stands for the total number of Weblog documents in the corpus of the Weblog search engine.

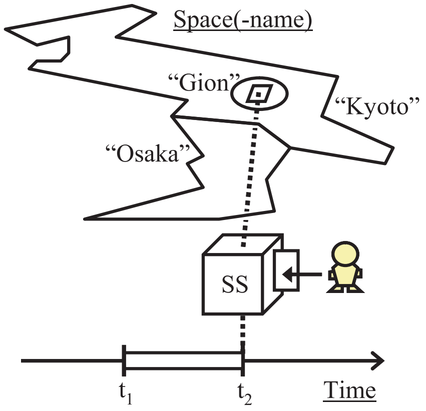

Fig. 4. Secure Space in "Gion" of "Kyoto"

IV. Experiment

First, we show two correlations on precipitation and traffic accidents to validate whether our Weblog sensor can sense phenomena in the real world. Next, we show several experimental results as examples of our Weblog sensor.

A. *Can the Weblog sense the Real World?*

Figure 5 shows the correlation of the weighting function weight(“Kyoto”, “Ame (rain)”) by our Weblog sensor with the amount of precipitation in “Kyoto” reported by the AMeDAS (Automated Meteorological Data Acquisition System) of JMA (Japan Meteorological Agency) [4] per day and month. The correlation coefficients per day and month are 0.624 and 0.925 respectively. In fact, we can see their strong correlation, especially per month. In this case, our Weblog sensor has enough strong correlation with the real world. If you want our Weblog sensor to alert yourself when weight(“Kyoto”, “Ame”) > 1.8, it is probable that you can protect yourself from the rain.

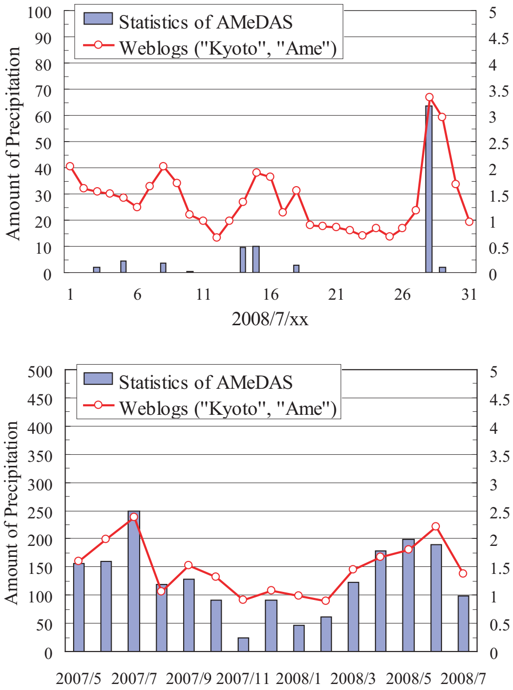

Fig. 5. Correlation of the relatedness of words “Kyoto” and “Ame (rain)” in the Weblog with the (real) amount of precipitation in “Kyoto” by the AMeDAS of JMA per day and month

Figure 6 shows the correlation of the weighting function weight(“Kyoto”, “Kotsu-Jiko (traffic accident)”) by our Weblog sensor with the number of traffic accidents in “Kyoto” reported by the Kyoto Police [5] per day and month. The correlation coefficients per day and month are 0.290 and 0.123 respectively. In fact, we cannot see their strong correlation, especially per month. In this case, our Weblog sensor has weakish correlation with the real world. Even if you want our Weblog sensor to alert yourself when weight(“Kyoto”, “Kotsu-Jiko”) > 1.8, it is doubtful that you can protect yourself from traffic accidents.

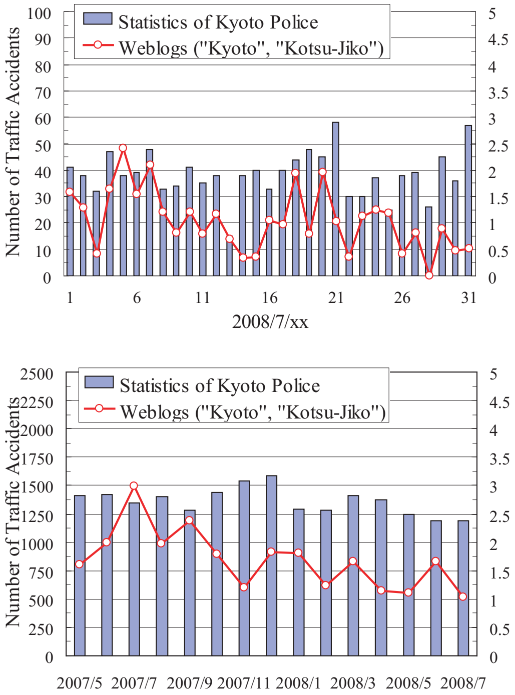

Fig. 6. Correlation of the relatedness of words “Kyoto” and “Kotsu-Jiko (traffic accident)” in the Weblog with the (real) number of traffic accidents in “Kyoto” by the Kyoto Police per day and month

B. *Examples of our Weblog Sensor*

Figures 7, 8, 9, and 10 in the next page show the degrees of “Yuutsu (dismal)”, “Konzatsu (crowd)”, “Kiken (danger)”, and “Hanzai (crime)” respectively in Secure Spaces named by “Kyoto”, “Gion”, and “Osaka” per day and month as examples of our Weblog sensor.

These spaces are almost not dismal, while they are almost crowded. Therefore, we cannot see the correlation of the relatedness of “Yuutsu” and them with the relatedness of “Konzatsu” and them in the Weblog. However, the degrees of “Konzatsu” for them by our Weblog sensor seem to reflect the real degrees of “Yuutsu” for them to some extent. Note that the Gion-Matsuri (Gion Festival) which is one of the most famous festivals in Japan takes place in July annually around Gion in Kyoto. In fact, within a few days before its climax, the Yamaboko-Junko, on July 17th, there are terribly crowded and those who hate crowded spaces would become dismal.

“Kyoto” and “Osaka” (which are broader areas) are almost dangerous and flooded with crimes, while “Gion” (which is a narrower area) is almost not so. Therefore, we can see the correlation of the relatedness of “Kiken” and them with the relatedness of “Hanzai” and them in the Weblog. Note that the degrees of “Hanzai” are more up-and-down than the degrees of “Kiken” depending on time, not only per day but also per month. And that crime-infested spaces imply dangerous spaces. Therefore, the degrees of “Hanzai” by our Weblog sensor might be better than the degrees of “Kiken” by our Weblog sensor as an approximation to the real degrees of “Kiken” for accee decision-making in Secure Spaces.

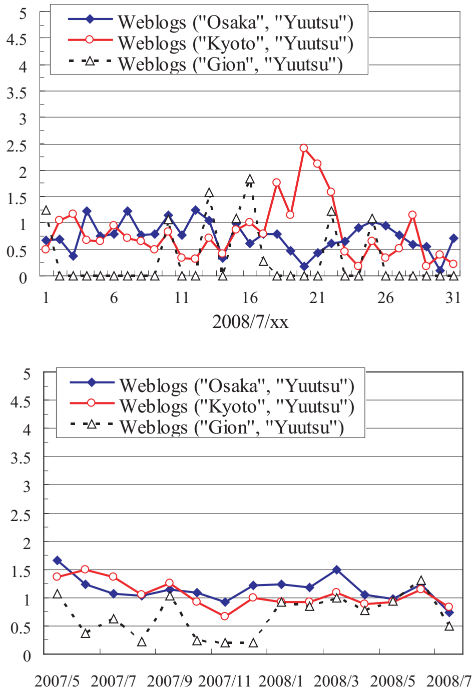

Fig. 7. Degrees of "Yuutsu (dismal)" in a Secure Space named by "Kyoto", "Gion", and "Osaka" by our Weblog sensor per day and month

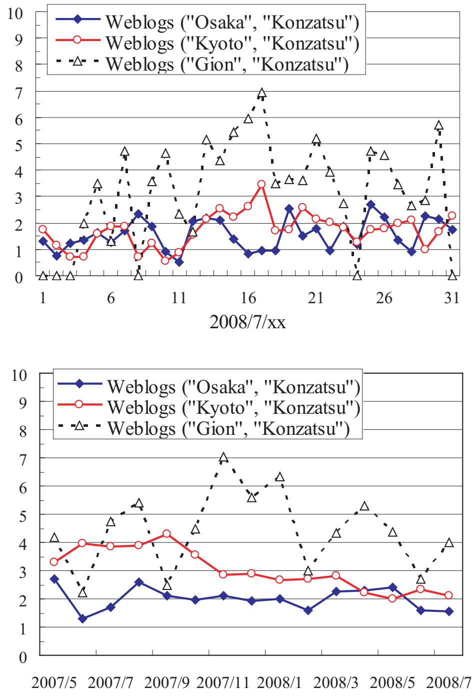

Fig. 8. Degrees of "Konzatsu (crowd)" in a Secure Space named by "Kyoto", "Gion", and "Osaka" by our Weblog sensor per day and month

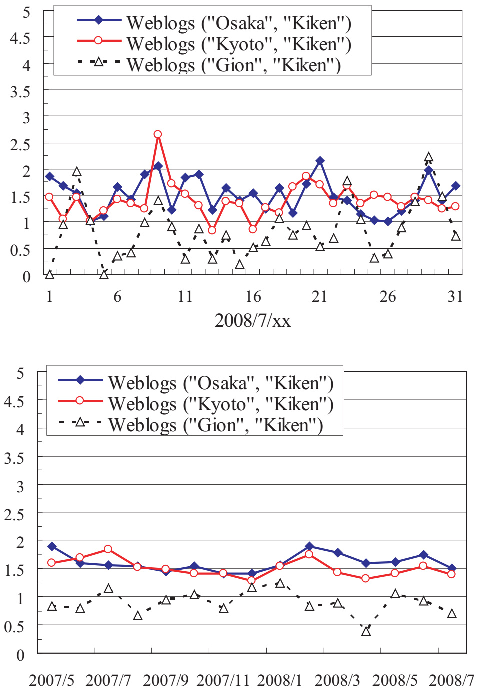

Fig. 9. Degrees of "Kiken (danger)" in a Secure Space named by "Kyoto", "Gion", and "Osaka" by our Weblog sensor per day and month

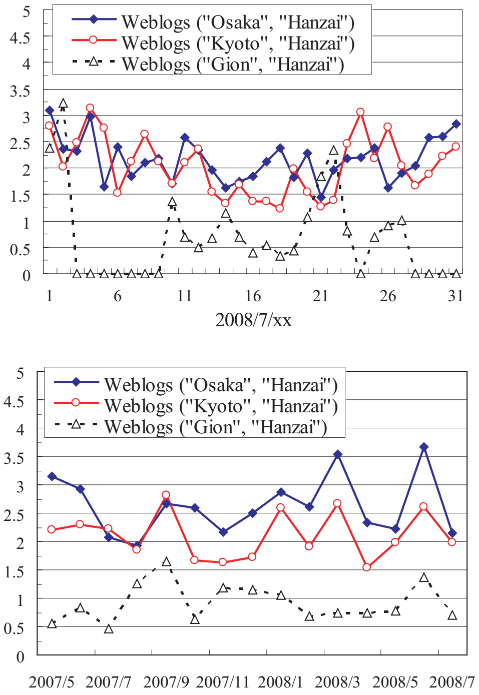

Fig. 10. Degrees of "Hanzai (crime)" in a Secure Space named by "Kyoto", "Gion", and "Osaka" by our Weblog sensor per day and month

V. Related Work

Our work in this paper and some previous ones [1], [2] are related to the research field of information security, especially authentication and access control, for Smart Spaces [6], [7]. They are often physically isolated environments such as rooms or buildings, which heterogeneous computing resources such as user input/output devices, various (real) sensors, actuators or communication apparatus are embedded in and which provide advanced services for their visitors by cooperating with their computational resources. They might have the capabilities to observe the real world, to interpret the observation, to perform reasoning on the interpretation, and then to perform some actions based on the reasoning. They are also called Active Spaces [8], Interactive Workspaces [9], Aware Homes [10], EasyLiving [11] and so forth. In [12], utilizing a room with two chambers for entry and exit, allows us to figure out those who enter and exit the room, that is, to identify who is in the room. When a user requests to enter a secured room in which an embedded display is outputting a virtual information resource that she does not have right to access, the room first revokes the output session and then grants her to enter itself. The secured room always denies nobody to enter itself unlike our entry control model. The other works on information access control for Smart Spaces include iSecurity which is a security framework for iRoom as an Interactive Workspace [9], a security architecture using Environment Roles for Aware Homes [10], a role-based context-aware access control mechanism within Hyperglue for Intelligent Environments [13].

Our work in this paper is also related to the research field of Web mining, especially spatially and temporally. Tezuka et al. [14] proposed a time-specific regional web search engine that enables users to retrieve both spatially and temporally restricted information from the Web. Kurashima et al. [15] proposed a system that extracts association rules between locations, time periods, and types of experiences as visitors' experiences from Weblog entries.

VI. Conclusion

As public spaces are made smarter by various information communication technologies, we might unexpectedly enter the public spaces that have our unfavorable characteristics (e.g., degrees of dismal and danger) and/or our unwanted information resources. Therefore, we have aimed at making Smart Spaces always secure in the real world and defined Secure Spaces in our previous works. In this paper, we proposed a method to extract information for making access or entry decisions in Secure Spaces from very large text corpora such as the Web, especially the Weblog, and also improved our previous architecture of Secure Spaces and mechanism of space entry control by adding the concept of the Weblog sensor, in order to enables users to more flexibly specify their access policies by keyword-based expressions. Several experimental results showed the potential of the Weblog sensor for access decision-making in Secure Spaces to some extent.

Acknowledgment

This work was supported in part by (1) a Grant-in-Aid for JSPS (Japan Society for the Promotion of Science) Fellows (Project Leader: Shun Hattori, Grant#: 1955301); (2) a MEXT (Ministry of Education, Culture, Sports, Science and Technology of Japan) Grant-in-Aid for Scientific Research on Priority Areas: "Cyber Infrastructure for the Information-explosion Era," Planning Research on "Contents Fusion and Seamless Search for Information Explosion" (Project Leader: Katsumi Tanaka, A01-00-02, Grant#: 18049041); and (3) the Kyoto University Global COE (Center of Excellence) Program: "Informatics Education and Research Center for Knowledge-Circulating Society" (Program Leader: Katsumi Tanaka).

References

[1] Shun Hattori and Katusmi Tanaka, *Secure Spaces: Protecting Freedom of Information Access in Public Places*, Proceedings of the 5th International Conference on Smart Homes and Health Telematics (ICOST'07), LNCS vol.4541, pp.99–109 (June 2007).

[2] Shun Hattori and Katsumi Tanaka, *Towards Building Secure Smart Spaces for Information Security in the Physical World*, Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), vol.11, no.8, pp.1023–1029 (September 2007).

[3] Yahoo! Blog Search, http://blog-search.yahoo.co.jp/ (2008).

[4] Japan Meteorological Agency, http://www.jma.go.jp/jp/amedas/ (2008).

[5] Kyoto Police, http://www.pref.kyoto.jp/fukei/ (2008).

[6] Lynne Rosenthal and Vincent: *NIST Smart Space, Pervasive Computing Initiative*, Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'00), pp.6–11 (June 2000).

[7] Jalal Al-Muhtadi, Anand Ranganathan, Roy Campbell, and M. Dennis Mickunas, *Cerberus: A Context-aware Security Scheme for Smart Spaces*, Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom'03), pp.489–496 (March 2003).

[8] Geetanjali Sampemane, Prasad Naldurg, and Roy H. Campbell, *Access Control for Active Spaces*, Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC'02), pp.343–352 (December 2002).

[9] Yee Jiun Song, Wendy Tobagus, Der Yao Leong, Brad Johanson, and Armando Fox, *Security: A Security Framework for Interactive Workspaces*, Technical Report, Stanford University (September 2003).

[10] Michael J. Covington, Wende Long, Srividhya Srinivasan, Anind K. Dey, Mustaque Ahamad, and Gregory D. Abowd, *Securing Context-aware Applications Using Environment Roles*, Proceedings of the 6th ACM Symposium on Access Control Models and Technologies (SACMAT'01), pp.10–20 (May 2001).

[11] Barry Brumitt, Brian Meyers, John Krumm, Amanda Kern and Steven Shafer, *EasyLiving: Technologies for Intelligent Environments*, Proceedings of the 2nd International Symposium on Handheld and Ubiquitous Computing (HUC'00), LNCS Vol.1927, pp.12–29 (September 2000).

[12] Patrick G. McLean, *A Secure Pervasive Environment*, Proceedings of the Australasian Information Security Workshop (AISW'03) in conjunction with the Australasian Computer Science Week 2003 (ACSW'03), vol.21, pp.67–75 (January 2003).

[13] Buddhika Kottahachchi and Robert Laddaga, *Access Controls for Intelligent Environments*, Proceedings of the 4th Annual International Conference on Intelligent Systems Design and Applications (ISDA'04) (August 2004).

[14] Taro Tezuka and Katsumi Tanaka, *Temporal and Spatial Attribute Extraction from Web Documents and Time-Specific Regional Web Search System*, Proceedings of the 4th International Workshop on Web and Wireless Geographical Information System (W2GIS'04), LNCS vol.3428, pp.14–25 (November 2005).

[15] Takeshi Kurashima, Taro Tezuka, and Katsumi Tanaka, *Mining and Visualizing Local Experiences from Blog Entries*, Proceedings of the 17th International Conference on Database and Expert Systems Applications(DEXA'06), LNCS vol.4080, pp.213–222 (September 2006).