DIGITAL TYPHOON: NEAR REAL-TIME AGGREGATION, RECOMBINATION AND DELIVERY OF TYPHOON-RELATED INFORMATION

Asanobu KITAMOTO
National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
kitamoto@nii.ac.jp

Abstract

This paper introduces the *Digital Typhoon* project, whose goal is to establish the study of meteoinformatics and to develop emergency information systems in the domain of typhoon-related information. We first propose the models of emergency information systems and uses "flat-source model" as the model of emergency information systems, and explain the concepts of information aggregation, recombination, and delivery, and describe how we used these concepts in the implementation of our web site named *Digital Typhoon* [1]. The basic principle of information recombination is the relationship that data plus context equals information, and we introduce the notion of context generation with real examples. Next information aggregation deals with content management systems for aggregating information comparing Weblog and Wiki, and we hypothesize that the effective emergency information aggregation systems need to have the hub of information, following the proposal of our information aggregation system using weblog trackbacks. Finally information delivery describes information needs and delivery tools for *Digital Typhoon*, which received more than 13 million page views since July 2003.

Keywords

Typhoon, Information aggregation, Information recombination, Information delivery, Context generation, Emergency information systems, Information hub.

1. Introduction

The year 2004 may be remembered as the year of disasters. Firstly, Japan suffered from the unprecedented number of ten typhoons that made landfall. Because many of those typhoons kept their intensity just before making landfall, heavy rainfall and strong winds resulted in the worst number of casualties in these twenty years. Secondly, the impact of tropical cyclones (typhoons and hurricanes) was huge not only in Japan but also in other regions of the world, including the series of hurricane strikes in the Caribbean Ocean and the United States, and the large number of casualties in Philippines due to four strikes of tropical cyclones. Finally, the year ended with the tragedy of earthquakes, including the Mid-Niigata Prefecture Earthquake in Japan on October, and the Indian Ocean Earthquake on December.

These disasters raised interests in information systems that work effective in emergency events. Today disaster management systems have already been designed and operated in various regions of the world, but they are usually designed in the framework of geographic information systems (GIS) and disaster assessment. We claim that we need something like emergency information systems specially designed by modeling the stream of information in emergency events. Geographical location is one of the most important information, but other types of data such as the stream of text, image, and video should be combined together to have a local and global view of emergency events. These ideas motivated us to propose the models of emergency information systems and discuss required concepts and technologies from the standpoint of informatics, which is the author's background.

The uniqueness of this paper comes from our experience with typhoon-related emergency information systems. Our experience has been accumulating through our project web site, *Digital Typhoon: Typhoon Images and Information* [1], which is designed to respond quickly to emergency events caused by typhoons. Based on this experience, Section 2 of this paper introduces the *Digital Typhoon* project with reference to the models of emergency information systems. From Section 3 through Section 5 is the main part of this paper with discussion on three stages of information stream, namely information aggregation, information recombination, and information delivery. The effective combination of those ideas is the foundation of our project. Finally Section 6 concludes the paper.

2. *Digital Typhoon* Project

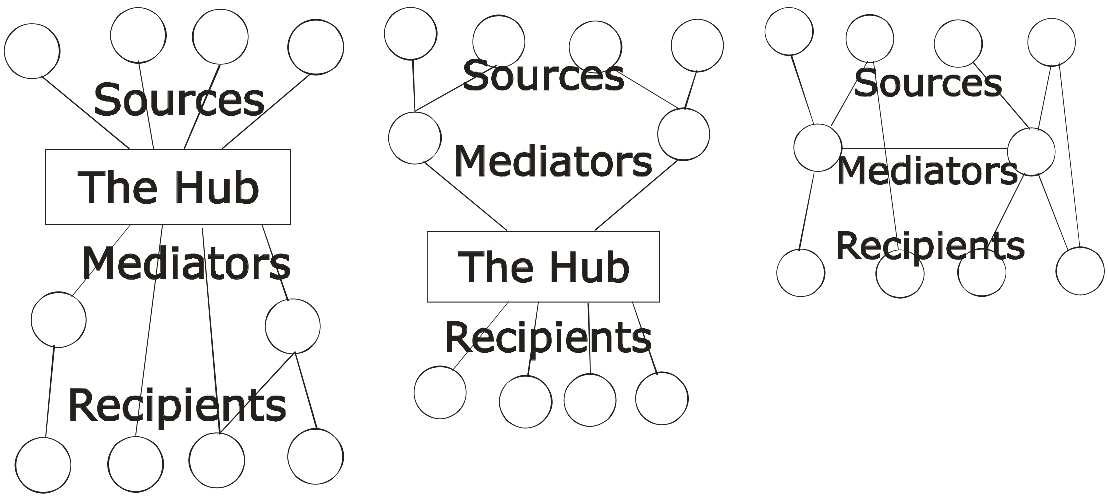
2.1 Background

The *Digital Typhoon* project started around 2000, and since then we worked on various fields ranging from informatics, remote sensing, and meteorology. We also started the web site *Digital Typhoon* in July 2003, and we have been providing near real-time information of typhoons to scientists and the general public in the world. The purpose of the project is two-fold.

1. Apply the meteoinformatics approach to typhoon image collection and study algorithms for efficient mining of rules and patterns from the huge amount of data.
2. Develop emergency information systems in the domain of typhoon-related information learning from real experiences with the web site.

This paper briefly refers to Point 1 in Section 2.5, and mainly discusses Point 2 in relation to *Digital Typhoon* project. Emergency information systems

2.2 Three Models of Emergency Information Systems



(a) Flat-Source Model (b) Structured-Source Model (c) Hub-less Model

Figure 1. The model of emergency information systems. A node represents a web site or a web page, and an edge represents relationship between nodes. In most cases an edge is a directed edge or a bidirectional edge, but we omit the direction for simplicity.

Figure 1 illustrates three models of emergency information systems. These models consist of four elements of the model; namely sources, the hub, mediators and recipients. The ultimate goal of emergency information systems is to propagate information from sources to recipients. To be more precise, the goal is to deliver the best emergency information from the best sources on requests from any recipients. How to design such information systems is the big challenge of research.

In the first place, sources are places where direct information or raw data is generated. Humans may produce live reports from the site, or sensors may produce monitoring results. They are frontlines to individual emergency events, and the stream of information or data generated by their eyes propagates through the Internet for wider people. The hub is the place where local information is aggregated and organized to have the global view of emergency events. Then mediators are people who collect information that they think is interesting or important, and contribute to propagating information to small or moderate number of people. Finally the recipients are people who just receive information and do not propagate information to others.

Depending on the type of information aggregation, the model can be further divided into two types, namely flat-source model and structured-source model. In the flat-source model, the hub aggregates streams of information from all sources. In the structured-source model, sources are structured so that filtering and editing of information can be applied in the middle of information aggregation. This structure is usually hierarchically organized to match the hierarchy of organization in governments or companies. The last model, the hub-less model, is a distributed system without the hub. Sources and recipients are generally linked in a peer-to-peer (P2P) style, but some of the recipients may voluntarily work as mediators to collect information based on personal motivation.

As three models illustrate, we assume that the presence of the hub and its role in the information stream is the important distinction of emergency information systems. The concept of information hub will be used throughout the paper, because we hypothesize that the presence of information hub at various levels is the important feature of emergency information systems. In the following, we briefly explain three models in more detail.

Flat-Source Model: It is unique in the aggregation of information. It focuses on the aggregation of information directly from sources, and the deliver of information after the appropriate organization of aggregated information. This model is similar to mass media, but we focus more on the combination of public/personal and direct/indirect information to provide multiple viewpoints for interpretation. We focus on this model in this paper.

Structured-Source Model: It is a typical structure in governmental emergency information systems whose hierarchical structure follows the hierarchical organization of governments. Hence this model is most efficient and can keep the quality of information. If it is combined with broadcasting mechanisms for information delivery, this model has an advantage in public awareness. The disadvantage is that it lacks the variety of information because personal information is usually filtered out in the hierarchy of information aggregation

Hub-less Model: It is gaining more interest recently today because it naturally matches with our notion that emergency information systems should not be a centralized system to eliminate a vulnerable element like the hub. This argument is true at least on the physical layer of the network, as shown in the success of the Internet. This distributed system is also effective from load-balancing point of view.

We argue in this paper that if we focus on information stream on the logical level, purely distributed systems are not effective for emergency information systems. Successful emergency information systems should have the hub for the aggregation, recombination, and delivery of information. This paper especially selects the flat-source model, and omits the structured source model because we think this is an already established model.

2.3 Three Stages of Information Stream

Three models of emergency information systems mainly concern different relationships between four elements along information stream. Logically this process can be decomposed into three stages, namely aggregation, recombination, and delivery. These stages are mainly performed around the hub, but it may be performed by other elements.

Information aggregation: It is to collect information for the hub from many sources. This process involves both push and pull of information. The push of information refers to the process that sources send information to the hub that they do have information. This type of aggregation mechanism has not been popular until recently, but the development of weblogs and the standardization of pushing information from weblogs, push is now a plausible solution for information aggregation. Another type, the pull of information, is more traditional which is usually implemented as a search engine crawler. This does not require the involvement of information sources, so the automatic aggregation of information is possible with this technique. This stage will be explained in Section 4.

Information recombination: It is to recombine data to make various meanings emerge from data. Just as the genetic recombination emerges new life from existing pairs of genome, information recombination emerges new meaning from existing sets of data. We assume that we discover new interpretation of data when data is put into appropriate contexts. So recombination is actually performed as interactions with databases, and the challenge here is to support various operators for creating interesting contexts from databases. This stage will be introduced in Section 3.

Information delivery: It is to deliver information to recipients via appropriate routes and timings. We can take advantage of various delivery paths such as web sites, mobile phones, syndication services and so on. The internet has many alternatives compared to mass media such as TV and radio, but because of that we need to combine various routes effectively to cope with urgent information needs during emergency events. This stage will be explained in Section 5.

2.4 Four Facets in *Digital Typhoon*

Digital Typhoon was originally designed to be the demonstrative web site of the project, but later evolved into the platform to test emergency information systems in the domain of typhoon-related information. It has four facets that can be represented in two dimensions as shown in Table 1.

	Direct (Primary)	Indirect (Secondary)
Public	Satellite images / AMeDAS [1]	Typhoon News Topics [2]
Personal	Eyes on the Typhoon [3]	Typhoon News Weblog [4]

Table 1. The classification of basic facets by two dimensions and respective web sites at *Digital Typhoon.*

The two dimensions represent the direct/indirect dimension and the personal/public dimension. The first dimension is related to the source of information. Direct information is generated from sensors or persons directly observing the events with their "eyes." On the other hand, indirect information is generated mainly from persons who combine direct information from multiple sources, filter irrelevant information, and edit relevant information in a concise manner. Usually direct information is published from many sources, while indirect information is published from limited number of sources.

The second dimension is related to the viewpoint of information. Personal information is described from a person's viewpoint that may be subjective but is expected to be consistent. On the other hand, public information is described to be objective and independent of individual viewpoints. Public information is more useful for understanding the events, while personal information is suitable for "feeling" the events.

2.4.1 Satellite images / AMeDAS (direct + public)

This facet deals with public data that come from various sensors or meteorological observations. Here public means that information does not depend on the interpretation of personal viewpoints (such as raw data from sensors), or information is about "facts." Hence this information is in general most reliable and serves as the basis of decision making for experts. However, this type of information is not enough. Firstly raw information is not suitable for non-experts who lack skills for appropriate interpretations. Secondly sensors and facts do not capture the entirety of emergency events. Usually one person or one organization can manage only a part of emergency information being generated massively concurrent, and one needs to depend on indirect information from reliable sources outside.

2.4.2 Typhoon News Topics (indirect + public)

Typhoon News Topics is the automatic summary of Japanese news articles distributed from mass media. News articles are first crawled and then text-matched to pick up news articles that describe each typhoon. To the luck of us, a typhoon is always referred by the standard name as "Typhoon No.23" (English translation of the Japanese expression), using the number of a typhoon, so the detection of typhoon news is a trivial task as long as we assume that the number always represents the number of the current year. Then we apply the morphological analysis of Japanese language and frequency of terms is calculated for choosing relevant terms from the corpus of news articles. These terms are summarized for each typhoon to highlight events, places and topics, and they are also used as keys to list news articles that contain the term. More advanced natural language processing should solve the problem of ambiguity in typhoon numbers. The number is recycled every year, so we have many typhoons with the same number. The resolution of the typhoon number requires the high level of natural language understanding, so it is left for future work.

2.4.3 Typhoon News Weblog (indirect + personal)

Typhoon News Weblog is the weblog-style narrative of typhoon-related information. The sources of information are satellite images and other meteorological observations that are archived locally in *Digital Typhoon*, or that are accessible on the Internet. Other sources include news aggregation services such as Yahoo! Japan News and Google News. The advances of Internet reduced the cost of, and increased the speed of, the activities of information searching and aggregation, and this is in favor of summarizing what is happening in the world. This facet belongs to the indirect + personal category, so the author should try, at least, to check the quality of information and arrange information as fair and consistent as possible.

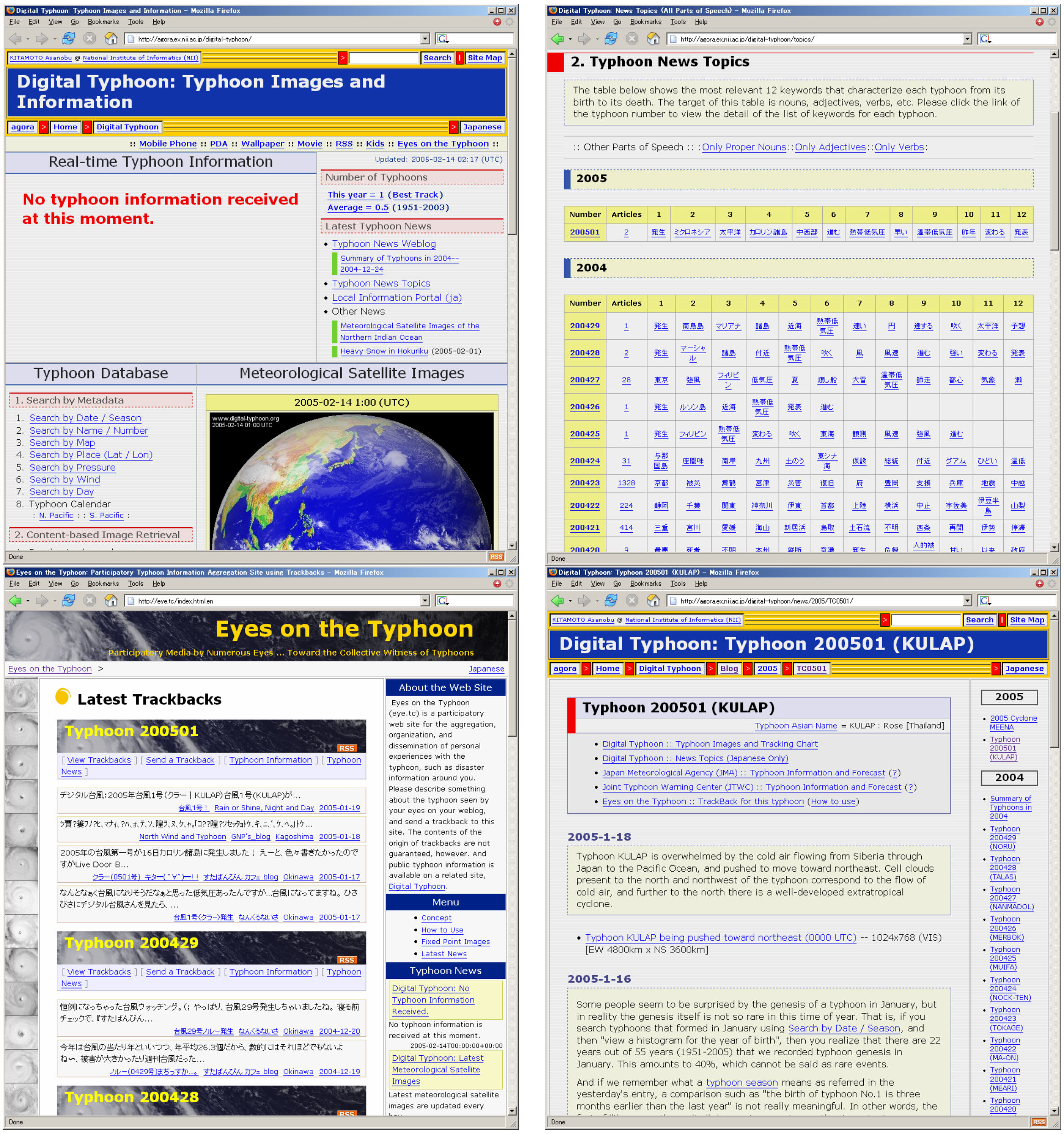


Figure 2. *Digital Typhoon* web pages. Following Table 1, web pages are arranged clockwise from the top left panel as *Digital Typhoon* (Top Page) [1], *Typhoon News Topics* [2], *Typhoon News Weblog* [3], *Eyes on the Typhoon* [4].

2.4.4 Eyes on the Typhoon (direct + personal)

Eyes on the Typhoon is designed for typhoon-related information aggregation from the general public. On the occurrence of emergency events, we need to have numerous eyes to record massively concurrent sub events at many places. Hence we need to ask other people to participate in the aggregation of local information at the information hub. The fundamental problem of participatory media, as this approach is called, is to find a good balance between quality and variety. Mass media focuses on quality, while personal media focuses on variety. Quality itself may be achieved by limiting sources to persons who have good domain expertise, but this limitation leads to information of less variety. Hence the challenge is to aggregate information from various sources and to improve the quality of information delivery. In the spectrum of quality and variety, participatory media tries to keep variety while maximizing quality. Information may not be correct, but we assume that more information leads to better decision. In emergency events, even unreliable information may help decision making as long as the reliability of information is properly treated, and people with good media literacy is involved. In any case, however, the challenge of participatory media is to establish mechanisms to support the judgment of quality.

2.5 Meteoinformatics Approach

Meteoinformatics approach is the fundamental principle of *Digital Typhoon*. This approach represents the database paradigm of meteorology, in contrast to computational meteorology, which represents the simulation paradigm of meteorology. Meteoinformatics can be characterized as learning from the past and use results for the future. Similar approaches to meteoinformatics are known as "the method of analogues" in meteorology. It is a method for predicting the future by searching for similar patterns in the past and comparing temporal evolution of similar cases. This method, however, lost its popularity because fundamental limitation of analogue-based methods was shown as the chaotic nature of atmosphere. However, the advancement of large-scale database technology now makes the meteoinformatics approach more realistic than before, and computers can take part in the extraction of useful rules and patterns from the huge amount of past data.

The particular target of meteoinformatics is the typhoon pattern recognition problem. It is motivated by the fact that the current typhoon (and hurricane) analysis method is based on the pattern recognition of typhoon satellite images by meteorological experts. This method assumes that similar-looking typhoons have similar features, and evaluate the similarity of patterns based on the domain knowledge of typhoons. Our goal is to apply pattern recognition algorithms on large-scale image collections to find an alternative method for this recognition problem.

As the basis of the meteoinformatics approach, we created the large-scale image collection of typhoon satellite images [5]. The database contains typhoon images comprehensively for nearly ten years, and as of January 2005, the image collection contains more than 56,300 satellite images, including 44,500 typhoon images for the north western Pacific, and others for the southern Hemisphere tropical cyclones around Australia. The uniqueness of this image collection lies in the fact that each image is "well-framed" with careful registration and map projection. Other data collected for the database includes more than 1,400 tracks of typhoons, and 300 million records of AMeDAS (Automated Meteorological Data Acquisition System) ground observations.

3. Information Recombination

3.1 Background

Information is not alone. Information has meaning, and the meaning of information emerges among relationships with the meaning of other information. We call these relationships as context. Our fundamental equation of information can be represented as follows.

Data + Context (+ Interpretation) = Information

This model claims that information emerges when a data is interpreted in a context. If a data is the same, but a context is different, then information is different because it is a function of a context. We insist that the generation of appropriate contexts that suppose meaningful interpretations of data is a fundamental role of databases. For the generation of contexts, we should recombine data in various ways.

Roughly speaking, a context is a set of data in which a particular data is to be interpreted in relation to other elements in the set. Hence we should recombine the set of data to generate another set of data, just as genetic recombination does on the set of genes to produce a new personality.

Let us suppose a simple example. We measure the temperature and get the data that temperature is 0 degree. This is a scientific measurement and nothing is ambiguous. But suppose we are asked to answer the question as follows: is 0 degree warm or cold? This trivial question cannot be answered in general because it depends on the context. If 0 degree is measured in the summer of Tokyo, it is extremely cold, but if 0 degree is measured in the Antarctic, it is very warm even in summer. This difference of meaning is not only due to subjective beliefs but also the context that the data needs to be interpreted; in this case the historical record of temperature at the particular place. For non-experts, absolute value is difficult to interpret, so we need to show them a context in which they can compare one data with others.

Hence the meaning of data, like warm or cold, should be interpreted in an appropriate context. Experts can produce appropriate contexts in their mind for the proper interpretation of data because they have accumulated knowledge about the data. Non-experts, on the other hand, see absolute values without appropriate contexts for interpretation, so they sometimes reach biased interpretation of absolute values. For example, frequent events are seen as rare events, or vice versa. To avoid such misunderstandings of data, we need to show data in an appropriate context, and this idea is what we call context generation.

3.2 Context Generation

Context generation may be confused with context extraction, but they are the opposite concepts each other. In context generation, we give contexts to a (ordered) set of data, but in context extraction, contexts are extracted from a set of data. The former concerns the selection of a set of data from the population of data, and arrange them in an appropriate manner to assist humans (or possibly computers) compare data from various viewpoints. Compared to content extraction, content generation is close to information design so that people can notice the meaning of information. This however is different from information design because we have interest in the dynamic generation of context with repeated interactions with databases. In short, information recombination processes consist of repeated generation of contexts using various criteria.

For interactions with databases, we need to have a language for communication, or what is called a query language for databases. This language defines basic operations for information recombination and the algebra for the combination of basic operators. The most popular query language is SQL (Structured Query Language) which is the standard language for relational databases [6]. On top of SQL, we implemented a part of a new query language for information recombination. We assume arrangement and grouping as the two fundamentally important operators. Arrangement criteria are chosen so that related data gets closer than otherwise, but the criterion may be a random shuffling, and all we assume is that data is arranged in some way. Grouping criteria are chosen so that related data is put into the same group, but the criterion may be a simple division, and all we assume is that data is grouped.

You may notice that similar operators are implemented in SQL as the ORDER BY operator and the GROUP BY operator. These operators, however, are implemented from practical needs on these operators, and they are not based on the original model of relational algebra. Moreover these operators have different functions from our expectation on these operators. Arrangement operators are not only about sorting but also about arranging data by any kinds of meaningful criteria. Grouping operators are not only about aggregation such as summation and average but also about extracting some of the data for further analysis. Thus those two operators have similar semantics but in fact are different from our operators.

Other basic operators are winnowing and expansion. Winnowing refers to the selection of some of the elements from a set. For example, it is used for selecting top-N elements from a set, which is a typical operation with the database for ranking data. Finally expansion refers to expanding attributes of data by calculating some values for each data. An example is similarity between data and the query. That similarity will be used for further analysis but cannot be computed beforehand because it depends on the query.

3.3 Examples

We now explain the example of context generation using exemplar navigation processes at *Digital Typhoon*. Suppose we start from the place of Tokyo (35.68N, 139.68E). In the first place we search for typhoon images that passed near this place in a track view as Figure 3-(a). This figure shows a context in a track view whose target set consists of typhoons that passed near Tokyo. From this context we can say that most typhoons that came to Tokyo was formed off the east cost of Philippines. Then we can search for typhoon satellite images for the same set of data but show information in a different context as Figure 3-(b). Then we can easily compare cases to notice that most of cloud patterns extend to north, which shows the sign of transforming into an extra-tropical cyclone. Further clicking of an image performs content-based image retrieval to search for similar images [7] to the query as shown in Figure 3-(c). Then we realize that there are many typhoons that have similar appearance and study further on the similarity of cloud patterns, such as how similar central pressure is among the set of similar images.

On the other hand, we can look for the history of typhoons that passed near Tokyo as Figure 3-(d) in a magnified scaling. From this context it is easy to understand that most typhoons move from southwest to northeast direction around Tokyo and other directions are relatively rare. This set of typhoons is viewed from different contexts, in the histogram of central pressure of typhoons that was recorded at Tokyo as shown in Figure 3-(e). Now we know that 940hPa is the minimum, and we had a typhoon of 950hPa near Tokyo. Then we are interested in this typhoon of 200003 (KIROGI) and view the summary of this typhoon as in Figure 3-(f).

These navigation examples show that the database of typhoons provides many operators for context generation. The basic operator is to arrange data by similarity in terms of geographical locations or cloud patterns. The results can be illustrated also in several views. Another operator is to arrange data according to the distribution of data like histogram. This context clearly represents which value is rare and which value is normal. Hence this context gives information for evaluating the risk of the situation in reference to history. At *Digital Typhoon*, we follow the context generation principle and try to show data combined with other data so that people can compare and study a set of data to have the better interpretation of data.

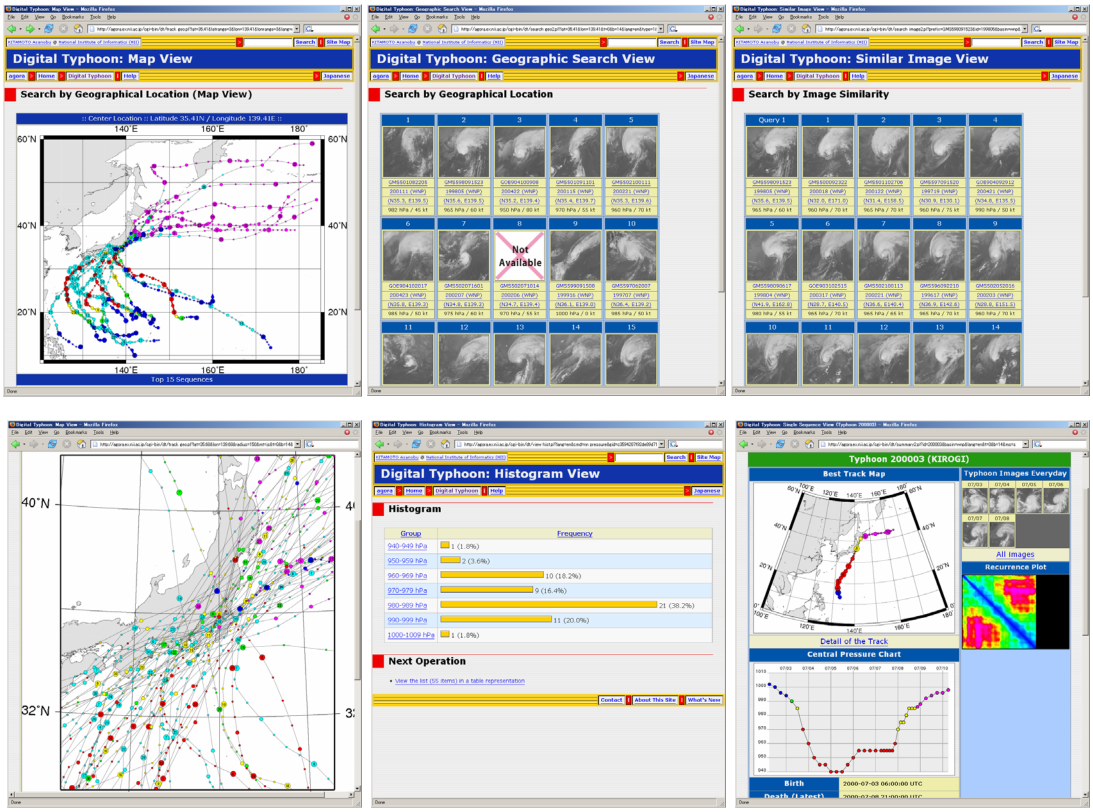


Figure 3. Examples of context generation. Panel (a)-(f) from the top left to the bottom right.

3.4 Future Directions

This information recombination system is implemented by a relational database engine (PostgreSQL), an image search engine (made by the project), and several scripts. At this moment, an image search engine can search for similar typhoon cloud patterns given a query image from the database. The theoretical foundation of the image search engine and its implementation is the current target of research. Traditional database engines focus on content-based retrieval, which uses the content of databases to find a set of data that match the query, such as keywords. What we should create is a database engine for context-based navigation, which supports repeated interaction with databases at the query language level to provide proper interpretations of data. The first version of this engine is already in operation at *Digital Typhoon*, and we are implementing the second version of this engine. Another continued effort was given to provide more contexts which are useful for the interpretation of emergency information in the domain of typhoons.

4. Information Aggregation

4.1 Experiences with Recent Natural Disasters

On the occurrence of emergency events, people can obtain local information within the extent of their observations, but global information can be obtained only through the aggregation of information from various sources and arrange them in a context. For this purpose, the Internet and the World Wide Web (WWW) proved its effectiveness as one of important platforms for emergency events. In Japan, the usage of the Internet and the WWW started just after the occurrence of Hanshin-Awaji Earthquake in 1995. Because of the great impact of the earthquake, we have seen many research products and practical systems to be started as disaster information systems. In the following, we evaluate the effectiveness of various systems in response to the Mid-Niigata Prefecture Earthquake in October 2004 and the Indian Ocean Earthquake in December 2004.

4.1.1 Safety Inquiry Service

The first type of systems is safety inquiry services. In the following we summarize the usage statistics of three representative systems for the Mid-Niigata Prefecture Earthquake.

Safety Inquiry Services	Registrations	Searches
NTT voice mail system [8]	About 113,000	About 236,000
NTT Docomo i-mode bulletin board [8]	About 88,000	About 116,000
I Am Alive (IAA) Alliance [9]	630	79,057

Table 2. Comparison of major safety inquiry services in Japan. The definition of registrations and searches may not be the same for each service. Statistics are temporary and approximate, and subject to change in the future.

Safety inquiry services are most necessary in the very beginning of emergency events. They are slowly increasing their popularity, but the number of users of the systems in Table 2 is much less than people staying in earthquake affected regions. This indicates that the usage of the system is still much lower than expected. Probably the biggest problem of them is the lack of interoperability across multiple regions, devices and organizations. It is often pointed out that we need the hub of safety information for sharing databases across regions, devices and companies, but this plan is yet to be realized. This problem is also important in the domain of typhoon-related information.

4.1.2 News Aggregation Service

Automatic news aggregation services such as Google News [10] showed their effectiveness as the hub of information in a global scale, especially as the hub of public and (semi-)direct information. Google News performs the frequent crawling of media companies' web sites and divides numerous news articles into the small number of clusters which are calculated automatically. The advantage of Google News is not only in the collection of worldwide news but also in providing access to local news which is difficult to reach otherwise. Usually local news is provided by local media companies and does not go outside of the region. This is also true in the scale of a country, and it is usually not easy to have access to local news of foreign countries. Google News, on the contrary, provides flat access to any news in the world, overturning conventional evaluation for the relevance of news. This is particularly important in terms of emergency information because we are usually short of local information which is important for the limited number of people. In the *Digital Typhoon* project, *Typhoon News Topics* plays a role similar to Google News. It performs crawling of news articles from mass media and applies the clustering of news articles by relevant keywords extracted automatically using natural language processing.

4.1.3 Weblog

Another tool that receives much attention recently is a content management tool called the weblog (or blog for short) [11]. The reason is that they are equipped with various functions that support non-experts to publish information. Based on the observation of weblogs in the context of emergency events, we suppose that weblogs can be classified into three types, following four elements introduced in Section 2.2.

Classification of Weblogs	Grades
Source Weblog	Very Good
Hub Weblog	Poor or Fair
Mediator Weblog	Fair or Good

Table 3. Classification of weblogs and their grades in the context of emergency information.

Source Weblog: They report local information from the viewpoint of people who stay on the site. It is often argued that people outside tend to have biased impression about emergency events, because mass media tend to concentrate on reporting about particular places that are most severely damaged. It is expected that source weblogs counterbalance biased impression by reporting local information directly from the site. The report needs not be too objective but is expected to be consistent and fair from the viewpoint of the person. Then the time-ordered entries of source weblogs become the invaluable records of what has happened in a microscopic scale.

Hub Weblog: They are intended to be the hub of information using weblogs as content management systems. The publishers of those weblogs do not have direct information in general, so they collect information from various sources, including other weblogs, and cite collected information in the order of time. However, we observed that hub weblogs performed poorly, especially in the beginning of emergency events, when many kinds of concurrent information streams are propagating across the Internet.

The problem here is the lack of hub within a weblog and the lack of consistency across weblog entries. The first problem refers to the equality of entries within a weblog. Every entry is treated equally as the independent and archival fragment of information, but emergency information is provisional and uncertain in general. Some information should be overwritten if more certain information or updated information is provided from information sources. Some entries or categories should be combined together to summarize available information at hand. Hence we claim that an entry is not the minimal information fragment in this case.

The second problem refers to the lack of consistency in hub weblogs because of the mixture of information from many sources. The advantage of weblogs lies in the consistency of entries based on the personal viewpoint, but the mixture of information with unknown quality destroys the coherency of weblog entries. After losing coherency of entries, the weblog is not an attractive solution as a content management system.

Mediator Weblog: They have only indirect information and help to propagate information about good sources and good hubs. The rapid propagation of good information can be compared to broadcasting in mass media, by which a large number of people quickly knows about good sources. Mediator weblogs may not contribute to add value to information, but they do contribute to open multiple paths to good sources.

This role, however, may often be misused. The copy-and-paste tradition and archive-oriented structure of weblogs resulted in the remaining of outdated information within citations that should be revised or deleted. Moreover, due to the high motivation of weblog publishers to make instant responses to events, incorrect information can be easily spread out through mediator weblogs or hub weblogs without careful scrutiny of information. In short, careful mediator weblogs are expected to play important roles for information propagation, while careless mediator weblogs may become the source of rumors and bring about the chaotic mixture of correct and incorrect information.

In *Digital Typhoon*, *Typhoon News Weblog* can be classified as either the source or the mediator weblog. As a source weblog, it publishes a kind of direct information created by analyzing or processing satellite images. As a mediator weblog it introduces good sources of information outside of *Digital Typhoon*. Another web site, *Eyes on the Typhoon,* takes the role of a hub weblog but in a different design because of the problems of the hub weblog. This site will be introduced in the latter part of this section.

4.1.4 Wiki

In contrast to the fail of the hub weblog, Wiki and its clones such as Wikipedia [12], has demonstrated potential as the hub of information. The advantage of Wiki is in the collaborative editing of information. The large number of editors collect information concurrently from various sources and what they think is important is picked up and put into the web site to be consistent with other information. Due to the lack of supervising editors, the quality of information at Wiki depends on the editors of individual articles, but it did work remarkably well for emergency events. Many types of information was well organized and frequently updated, and old information was overwritten when it became outdated. Adaptive editorial policy also enabled the dynamic splitting of one concept to multiple sub-concepts when the amount and extent of information exceeded the appropriate level of a single concept. As the hub of information, these flexible editorial tricks made Wiki-based systems more usable than the hub weblogs that stick to the rigid design of information.

4.2 URI-based Information Hub

Here we are especially interested in the success of Wiki and fundamental difference between Wiki and Weblog when they are used as the hub of information. We hypothesize that the advantage of Wiki is in the URI (Unique Resource Identifier)-based design of representing a term or a concept. The URI-based design allows people to identify appropriate URIs for contributing their information, and aggregating all related information into one place. On the contrary, weblogs do not have meaningful URIs for a term or a concept. They do have URIs for categories, but in general they are just the container of entries and do not serve as the hub of information. This entry-based design prevents the efficient aggregation and organization of information, resulted in redundant and dispersed collection of information.

As a result we follow the URI-based design. We first begin with the definition of URI for the hub of information, and then use the associated URLs for the so-called trackback from (source and mediator) weblogs. For the definition of URIs we employ what we call “location-embedded trackback URL” that encodes location information by the concatenation of location codes. We choose location codes so that one addressing scheme does not share the same code with other addressing schemes. Hence we can uniquely decode location information from a location embedded trackback URL without tagging of the codes, and associate that information with appropriate URI for the aggregation of information. There are two choices for addressing schemes.

Flat-Addressing Scheme: It uses uniform addressing scheme such as latitude and longitude. The advantage of this scheme is that system does not need much knowledge about addressing schemes, and it can be easily integrated with location-aware devices such as GPS (Global Positioning System). The disadvantage of this scheme is that the identification of location is not efficient without such devices. We can do it by the zooming in and out of the map, but it needs a lot of investment on using maps, which may not be required otherwise.

Hierarchical-Addressing Scheme: It depends on the existing organizational structure of the government or other addressing schemes such as the postal code. Then the location is expressed as the hierarchical (part-of) structure of codes. An example is nation-prefecture-city hierarchy in Japan. An advantage of this scheme is that people are familiar with the system and can easily navigate through the system of codes. Another advantage is that the level of detail along the hierarchy can be used to as a built-in mechanism to select the level of privacy. The pinging of information from their weblogs reveals information about the location of weblog publishers, so they choose appropriate privacy level at their preference. This kind of built-in mechanism for privacy is not available in the flat-addressing scheme.

We follow the hierarchical-addressing scheme and define a location-embedded trackback URL as follows.

http://eye.tc/trackback/ping/200423/1018430

This means that the trackback information should be sent to the hub of Typhoon No.23 of 2004 at the place encoded as 1018430. Then 1018430 can be decoded as the 7-digit postal code and lookup the table of postal codes to find the real place. In summary, we list other coding schemes as follows.

Regular Expression of Codes	Coding Scheme
¥w{2}	ISO-3166-1 Country Codes
¥d{2}	JIS-X0401 Prefecture Identification Code (Japan)
¥d{5}	JIS-X0402 Towns and Villages Code (Japan)
¥d{6}	Typhoon Number YYYYNN
¥d{7}	7-digit Postal Code (Japan)
¥d{8}	Date YYYYMMDD

Table 4. Addressing schemes and their codes classified by regular expression of code strings.'¥' means a meta character, 'w', an alphabet, and 'd', a digit. Finally {x} means the length x of a code string.

4.3 Information Recombination for Aggregated Information

(a) Distribution of trackbacks by prefecture for Typhoon 200423.

(b) The list of trackbacks for Typhoon 200423.

(c) The list of trackbacks for Typhoon 200423 from Okinawa.

Figure 4. Information recombination for aggregated information. Various types of contexts are provided at *Eyes on the Typhoon*.

Each trackback contains independent information from weblog publishers. Context generation in this case is simple: the name of the typhoon, the location of the source of trackbacks, and the combination of those, as shown in Figure 4 (a)-(c). For example, Figure 4-(a) shows the distribution of the source of trackbacks by prefecture is illustrated by a simple map which mimics the map of Japan. Out of 43 trackbacks received for Typhoon 200423, 32 trackbacks are pinged for location-embedded trackback URLs and other 11 trackbacks were pinged for the trackback URL that only specifies the name of the typhoon. With this simple interface, we can see that which regions are affected by the typhoon, and also how those regions are affected by jumping to original sources. We have no mechanisms to prove that those trackback pings were actually sent from respective places. Nevertheless this result shows that with the help of motivated people who want to broadcast local information to wider communities, this mechanism worked effective to match their needs. Other interface includes the view of trackbacks in a list form for each typhoon as Figure 4-(b), and the combination of the typhoon name and the prefecture as shown in Figure 4-(c). The latter interface allows us to "zoom in" to more localized information per town, village, and postal code.

4.4 Current Status and Future Directions

Eyes on the Typhoon has received 168 trackbacks from weblogs since June 2004. The maximum number of 43 trackbacks arrived for Typhoon 200423 (TOKAGE), and other typhoons also received a smaller number of trackbacks. The number of trackbacks for each typhoon is in general increasing as both the popularity of the web site and the number of weblog users increases simultaneously.

The contents of trackbacks are versatile, including live reports from the site which is under the strike of the typhoon, or the damage of typhoons after the typhoon passed the site. Other contents include the introduction of the site, the technology used, and their own analysis of the typhoon. Thus the site attracted many people who have interest in typhoons, and served as the hub of information.

Nevertheless the current status is far from complete. One obvious extension of this system is to accept information from mobile phones. They are the best devices on the site of disasters because of its mobility and battery, and communication facilities. It is also common to have a camera and a GPS, which produces valuable information as the description of emergency events. The usage of mobile phones in emergency events is itself not unique, but their use is usually restricted to an environment where users are limited to registered ones. To keep the quality of information in open systems is the fundamental challenge of this line of research.

To improve the quality, one possible solution is the introduction of metadata for describing emergency information. Metadata is the standard format for describing information, and can give more semantics to natural language text or numeric values. For example, the time and the place of the site can be described more correctly if everyone follows standard procedures to describe emergency information. It is also important to define technical terms and their agreed meanings. What is important in the argument of metadata, however, that metadata should be introduced along with good interface or tools for inputting metadata. Otherwise people feel that metadata input requires too much effort that does not match with merits. Hence the development of metadata and useful tools requires the large-scale deployment of activities for emergency information systems.

5. Information Delivery

5.1 Urgent Needs for Typhoon Information

The number of page views at *Digital Typhoon* is still increasing. The total number of page views in one and a half year since July 2003 amounts to about 13.3 million, including 3.8 million to the top page. These numbers clearly demonstrate urgent needs for typhoon information, but the statistics of information needs changes bursty depending on the location and direction of the typhoon. As Figure 5 shows, we observe the following statistics from the log files of the web server.

1. Information need increases bursty when a typhoon becomes active. The peak information need is roughly ten times larger than that of normal periods.
2. A closer look at the statistics show the page view increases as the typhoon draws near to Japan.

In fact, the number of page views exceeded 410,000 on October 2004 on the approaching of Typhoon 200423 (TOKAGE). In the first stage, only people who regularly visit this web site or subscribe RSS news feed can notice the birth of a typhoon. Then some users put this news on their web site / weblogs, and information is distributed slowly among people. Then if a typhoon draws near to Japan, weather forecast programs of mass media start the warning against the typhoon, and more people notice the presence of a typhoon. Finally if a typhoon is expected to make landfall, regular TV news and newspapers start special programs to deliver typhoon information and then information needs reach the maximum.

This kind of statistics may be unique to typhoons compared to other natural disasters or man-made disasters. The latter disasters cannot be predicted in advance, but nowadays the approaching of typhoons can be roughly predicted a few days in advance, and information needs grow steadily as the probability of landfall increases according to weather forecast, as illustrated in Figure 6.

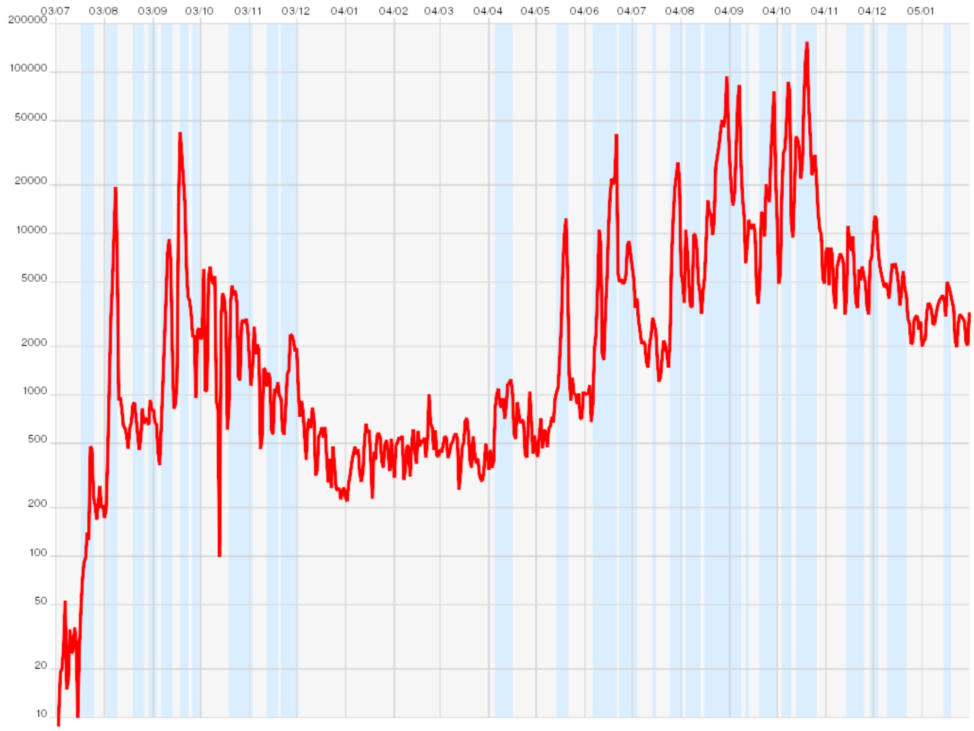


Figure 5. The daily number of page views to the top page of *Digital Typhoon* since July 2003. The colored bands represent the life of typhoons and note that page views increase during the presence of active typhoons. Several peaks correspond to the day when a typhoon made landfall on Japan. The peak page view is about 150,000 page views per day, which corresponds to the peak of Figure 6.

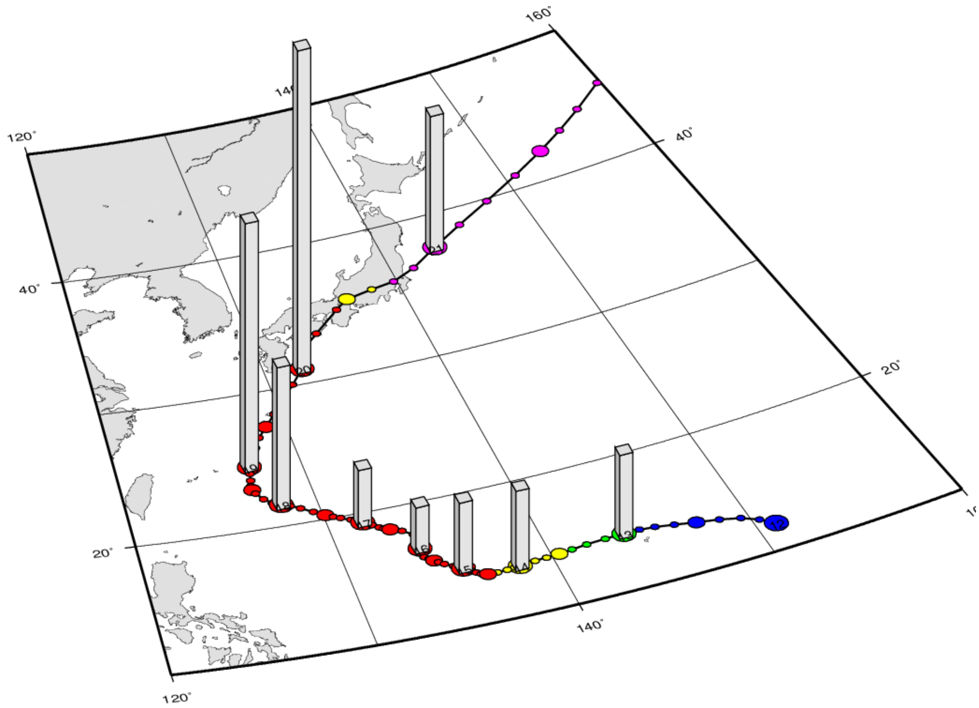


Figure. 6. The number of page views to the top page of *Digital Typhoon* during the life of Typhoon 200423 (TOKAGE). The number of page views is represented as the height of the bar at the location of the typhoon at 00 UTC of the day. The peak page view is about 150,000 page views per day, which

corresponds to the peak of Figure 5.

5.2 Information Delivery for Urgent Information Needs

The burst of information needs on the approaching of typhoons should be handled by introducing other information delivery mechanisms than browsing the web site. At this moment we have two other channels for the delivery of information.

One is a mobile phone interface for typhoon information. In emergency situations, people may not have personal computers with them, and the most probable device that they have for obtaining information is a mobile phone. Hence we provide a web page that can be browsed by a mobile phone to get the minimal set of information including satellite images. It is reported that, in emergency situations, a mobile phone is often one of the last devices they can use for collecting information, because other devices are more vulnerable to power failure or the destruction of infrastructure. Hence the mobile phone is the important device for information delivery.

Another interface for information delivery is a syndication service such as RSS (Rich Site Summary or Really Simple Syndication). *Digital Typhoon* has been providing the RSS service since September 2003, and it receives about 2000 daily accesses from the world. RSS contains the minimal information about typhoons, so people who do not need to know the detailed information can just check RSS and obtain information in a concise manner.

In the future, an important challenge is to build the distributed system of *Digital Typhoon*. On the logical level, the web site needs to be designed as the hub of information, but it does not mean that it should be built as the single central server in the physical level of the network. Japan Meteorological Agency, for example, introduced content delivery service that has multiple cache servers at major ISPs (Information Service Providers) in Japan, and the web site is now accessible even during the massive amount of accesses during the landfall of typhoons [13].

Another solution, P2P-based software such as *bittorrent*, was used for the Indian Ocean Earthquake in response to very high demand on the video shots of tsunami. This strategy proved to be effective from load-balancing point of view, but P2P-like system is not effective from information aggregation point of view. For example, weblog trackbacks are easily led to a mingled structure when weblog publishers are allowed to ping trackbacks without constraints. This strategy may be effective in normal occasions, and probably the ecology of weblogs naturally forms the hub of information as time goes. Emergency events, however, spawn urgent information needs, and we cannot wait until the hub of information naturally emerges. We need to prepare the hub of information before emergency events, or operate it any time regardless of the presence of emergency events. To cope with massive number of accesses to the hub, however, we need to realize the distributed system in the physical layer, while maintaining the hub in the logical layer.

6. Conclusion

This paper introduced the project *Digital Typhoon* and discussed basic principles in the designing of typhoon-related emergency information systems. We proposed the models of emergency information systems and identified important topics that need to be discussed. Three stages of information stream -- namely information recombination, information aggregation, and information delivery – should be supported by emergency information systems and four facets in *Digital Typhoon* help deal with various types of emergency information. As a result, *Digital Typhoon* has received more than 13 million page views and is emerging as the hub of information in the domain of typhoons. Although this activity is experimental in nature, we wan to contribute to the society by providing our skills in managing various types of emergency information effectively.

It is pointed out that we will see more natural disasters, at least meteorologically-caused disasters, due to the increase of severe weather resulted from climate change, or global warming. The scale of damage from natural disasters also tends to increase as the society gets more sophisticated to depend heavily on technology such as electricity and computers. Vulnerability to disasters should partially be reduced by emergency information systems, and we hope that our proposal contribute to this line of research.

An interesting research direction is the development of a system for incorporating the power of amateurs. As the Indian Ocean Earthquake shows, photographs, videos and narratives taken by ordinary people became the valuable records of the earthquake from various viewpoints. Massively concurrent recordings of emergency events were not realistic before, but in the 21st century we have ubiquitous sensors such as personal digital cameras, video cameras and mobile phones, and those recordings will be valuable materials for research and education in the future. This trend became apparent when the terrorism attack of September 11th in the United States was recorded by many people with their cameras and reports, most of which were made by amateur people and tourists. A system that serves as the hub of information from amateurs may become the valuable resource of information, but it must be equipped with filtering, editing, and organizing of information for the efficient delivery of information and the prevention of rumors. Now the Internet and the WWW became an indispensable media to deliver official emergency information from the government or other special organizations. The next step is to collect the power of people to respond personally to emergency events.

Acknowledgment

Satellite images used in this research is GMS-5 and GOES-9 satellite images that are distributed either from Institute of Industrial Science, University of Tokyo, and Japan Meteorological Agency. Best track dataset of typhoons is also distributed from Japan Meteorological Agency. Some maps and images are drawn by Generic Mapping Tools (GMT) by University of Hawaii, and Blue Marble by NASA.

References

[1] Digital Typhoon, http://www.digital-typhoon.org/ .
[2] Typhoon News Topics, http://agora.ex.nii.ac.jp/digital-typhoon/topics/.
[3] Eyes on the Typhoon, http://eye.tc/
[4] Typhoon News Weblog, http://agora.ex.nii.ac.jp/digital-typhoon/news/.
[5] A. KITAMOTO, "Spatio-temporal Data Mining for Typhoon Image Collection", Journal of Intelligent Information Systems, Vol. 19, No. 1, pp. 25-41, 2002.
[6] C.J. Date, "An Introduction to Database Systems", Addison-Wesley, 2000.
[7] A. KITAMOTO, "IMET: Image Mining Environment for Typhoon Analysis and Prediction", Multimedia Mining, Djeraba, C. (Ed.), pp. 7-24, Kluwer Academic Publishers, 2002.
[8] ITmedia Mobile, http://www.itmedia.co.jp/mobile/articles/0411/01/news093.html, 2004 (in Japanese).
[9] I Am Alive (IAA) Alliance, http://www.iaa-alliance.net/, Statistics is from personal communication, 2005.
[10] Google News, http://news.google.com/ .
[11] Rebecca Blood, "The Weblog Handbook," Perseus Books, 2002.
[12] Wikipedia, http://www.wikipedia.org/ .
[13] Japan Meteoroloical Agency, http://www.jma.go.jp/JMA_HP/jma/press/0409/16b/hpaccess.pdf, 2004 (in Japanese).