Overview of MuST at the NTCIR-7 Workshop – Challenges to Multi-modal Summarization for Trend Information –

Tsuneaki Kato
The University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
kato@boz.c.u-tokyo.ac.jp

Mitsunori Matsushita
Kansai University
2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka, 569-1095 Japan
mat@res.kutc.kansai-u.ac.jp

Abstract

The Workshop on Multi-modal Summarization for Trend information (MuST) has been encouraging researches on summarizing trends, which are an important first step of researches on technologies for supporting interactive and explorative information utilization. While MuST in the previous cycles provided the participants with a framework for encouraging researches by sharing the same research resources and by addressing a common theme in a broad sense, MuST at NTCIR-7 had a new aspect of an evaluation workshop by picking up some themes that many participants have been tackling and organizing those into evaluation subtasks. These two aspects of the workshop have helped to activate and push forward the researches: one shows a wide range of possibilities and visions in this field, while the other confirms the state of accomplishment. In addition, shared resources have been developed through the workshop. These activities are encouraging researches on multi-modal summarization of trend information.

1 Introduction

Information access is becoming increasingly interactive and explorative. Information gathering is no longer just a one-shot interaction in which a user describes his/her interest precisely and in return obtains just enough relevant information. Rather, the process is now a continuous one in which users browse information gathered according to their vague interest, and then find interesting parts on which more detailed information is required. Through this process, by elaborating areas of interest, users interactively focus on some information. Another crucial trend of information access is that information is presented not only as text, but by several types of media.

For supporting such interactive and explorative information utilization, technologies with at least the following two characteristics are required. Firstly, a framework is needed that seamlessly supports users throughout the information access process, from browsing an outline or summary, to subsequent elaborating and narrowing down, and to obtaining proper information. Secondly, linguistic and non-linguistic information should be cooperatively employed in this process: information handled should not be limited to text, but should include non-linguistic information such as numerical values, and non-linguistic modes should also be used, including multi-modal presentation and information visualization.

The Workshop on Multi-modal Summarization for Trend information (MuST) has been fostering researches on summarizing trends, which are an important first step of developing those technologies. Trends are the first answers to users' questions such as "How has the game machine industry performed since 2006?," "How have oil and gasoline prices changed this year?" and "How bad were the typhoons last year?" These can be considered a summary of all the information that users are interested in and a starting point for interactive and explorative information access.

Information composing trends and the process of compiling trends have several interesting features such as the following.

- It is necessary to compile information ranging over a given, considerable time period to obtain trends. Since it includes a lot of redundancy, such compilation needs to be synthetical and well organized.

- Trends usually contain summaries of non-linguistic information, that is, statistical information such as time-series data and geometrical data. Some statistics such as political party approval ratings and companies' shares of a given product are more complicated and have other dimensions. Each dimension can be an axis of representing those statistics and gives a different summarization method.

- Not only factual information such as reports on changes of statistical data, but also their interpretation, analysis of the causes, and forecasts of impacts are important and should be included in trends.

- Gathering information for trend compilation should cover not only several genres of textual information such as newspaper articles and blog pages, but also many styles of information outside texts such as numerical information on white papers.

Since compiling trends requires such a sophisticated process for handling complex, varied information, it is an important research subject for supporting interactive and explorative information access. And it is also expected that the research results can be applicable to other kinds of information. In this sense, the results of MuST can be deployed in various fields of information access, combined with technologies developed through other clusters of the NTCIR workshop.

This report summarizes the MuST workshop conducted at the NTCIR-7 workshop. First, its positioning and the specific framework of MuST in NTCIR-7, which consists of free subtasks and evaluation subtasks, are described. The progress of the current workshop is then reviewed. Following an explanation of three research resources constructed through the workshop, detailed reports on the free subtasks and evaluation subtasks are presented. Lastly, the contribution of the workshop is discussed and conclusions are presented.

2 The Positioning and Framework

MuST has been conducted since 2005 as a pilot task of the NTCIR workshop. It has been managed with a cycle of around one year. The second cycle was completed in March 2007. MuST in those cycles was designed to encourage cooperative and competitive studies on multimodal summarization for trend information. One purpose of the workshop was to promote discussion and community conformity, while another was to construct and accumulate resources such as tools and corpora. This would be encouraged by sharing the same research resource and by conducting research concerning a common theme in a broad sense. An objective and quantitative evaluation using a test set was outside the scope of those cycles of the workshop, which distinguished this workshop from other NTCIR tasks [4, 5, 6].

The major characteristic of the MuST workshop has been the use of a shared resource, which motivates participation in the workshop and substitutes for objective evaluation using a common test set in evaluation workshops. Although multimodal summarization for trend information tackled in MuST requires several component technologies dispersed in many research fields, which makes it difficult to construct a total system, by using a shared resource the participants can address their own subjects. The shared resource, *the MuST Data Set*, includes not only the materials to be processed, but also the corpus with annotations of intermediate results of compiling trends. Participants from different communities can discuss their interests with each other using the data set as common ground, and can see how their studies or their modules fit into the whole framework. Of course, researchers having the same interest can use the data set as a material for objective evaluation. Fostering researches through such interchanges was the objective of the first and second cycles of the MuST workshop.

Based on the course so far, MuST at NTCIR-7 has become two-fold. First, it follows the course so far, in which researches originated from each participant are encouraged by sharing the same research resource and the common theme in a broad sense. It is a workshop for deriving advantages of having a research community. These research themes are called *the free subtasks*. Second, by picking up some themes that many participants have been addressing in the previous cycles and organizing those into *evaluation subtasks*, MuST has grown into an evaluation workshop.

The organizers have conceived the following three evaluation subtasks, all of which are fundamental technologies for interchanging between textual information and numerical information or for aligning those two kinds of information, which in turn are the core of multi-modal summarization for trend information.

T2N subtask
: *Conversion from text to numerical data or numerical data extraction from texts,* in which systems are expected to extract time-series statistical information from a set of texts, that is, newspaper articles, and to compile it into a given format. This makes it possible to identify what time point and what value of a given statistic is discussed in those texts. The data extracted can be used also for visualizing the statistical changes according to the interests expressed in those texts.

N2T subtask
: *Conversion from numerical data to text or text*

Table 1. The Participants to MuST at NTCIR-7

Organization	Reference	Free	T2N
A.I. Lab., Dept. of Information Engineering, Faculty of Engineering, Mie University	[14]		Mie
Communication Design Lab., Dept. of Information Systems and Multimedia Design, School of Engineering, Tokyo Denki University	[16]	*	deNdai
Computational Linguistics Group, National Institute of Information and Communications Technology	[10]	*	
Dept. of Computer Science, Graduate School of Natural Science and Technology, Okayama University	[3]	*	
Dept. of Mathematics and Information Sciences, Graduate School of Science, Osaka Prefecture Univerisry	[15]	*	
Dept. of Information and Management Science, Otaru University of Commerce	[12]	*	
Faculty of System Design, Tokyo Metropolitan University	[13]	*	
Graduate School of Information Sciences, Hiroshima City University	[11]		HCU1,···, HCU4
Itoh Lab., Dept. of Information Sciences, Ochanomizu Univerisry	[2]	*	
Kobayashi Lab., Ochanomizu University	[8]	*	
Mori Lab., Graduate School of Environment and Information Sciences, Yokohama National University	[9]	*	Forst
Saito Lab., Dept. of Information and Computer Science, Keio University	[1]		Keio1, Keio2
Service Platforms Research Laboratories, NEC	[7]	*	

generation from numerical data, in which systems are expected to generate a paragraph-length text describing the changes of a given time-series statistical information represented in a table of the sequence of numerical values. Texts generated are linguistic summaries of numerical information.

ALN subtask
Alignment of textual information and time-series statistical data, in which systems are expected to extract statements on a given statistic in a given text set, and to link those to the corresponding portions of the table of the sequence of numerical values concerning the same time-series statistical information. Statements to be linked include not only quantitative expressions but also more qualitative expressions such as “peak” and “gradual decrease.” It is inter-media alignment, which allows systems to cooperatively use information in different media from different sources.

In addition to the proposal and elaboration of those subtasks, the organizers have planned to develop and deliver some resources in order to support researches on both free and evaluation subtasks. As well as the case in the previous cycles, sharing such resources is expected to make the workshop more attractive and effective. Those resources include, in addition to *the MuST data set* mentioned above, *a Change Expression Corpus*, which provides references for T2N and ALN tasks, and *a Visualization Platform*, which makes it easier to construct multi-modal presentation systems. These are explained in section 4.

3 Progress

MuST at NTCIR-7 was first announced in November 2007, at a tutorial meeting on NTCIR-7. The workshop started rather late compared to other clusters since it took longer to design a new MuST workshop and to decide how to carry it out. The call for participation closed at the end of 2007; MuST had 18 groups of participants, five of which later withdrew their participation. Table 1 shows the names of the active participants. A round-table meeting was held in January 2008, where the organizers explained the course and the direction of the new style of MuST workshop. An intermediate workshop meeting was held in March 2008, where the participants reported on the progress of their free subtasks and the organizers reported on the construction of the resource. Eleven reports were presented, which are available on the MuST homepage[1]. Through the discussions at this stage, it was decided to conduct only the T2N subtask in MuST at NTCIR-7, because the other evaluation subtasks were not sufficient to attract many participants.

The T2N subtask was carried out as follows. No dry run was conducted. For reference, the change expression corpus, which is explained later in section 4.3, was constructed and delivered, which covers five topics and fifteen statistics picked up from the existing MuST data set. A tentative version was delivered in May and the final version in June. The subjects of the formal run were delivered on July 16, and the deadline for submission was July 31; the period of the formal run experiment was around two weeks. The change

[1] http://must.c.u-tokyo.ac.jp

expression corpus on the formal run subjects and the numerical data on those statistics, which were gathered from other resources such as web pages, were delivered on August 1, soon after the end of the formal run. As the first evaluation, the precision and recall of submission results of T2N tasks were delivered on September 6.

4 Resources

4.1 The MuST Data Set

The MuST data set was designed and constructed with the objective of becoming a shared research resource for fostering the development of a system for compiling trend information and researches on their component technologies. It consists of materials used to compile trend information, intermediate results of the compilation process, and examples/references of components of multimedia presentations to be obtained on several topics, each of which corresponds to a user's query or information request. The materials are newspaper articles on selected topics, and the intermediate results of the compilation process, which are the most important part of the data set, are annotated on those articles. Since those annotations identify the names and values of statistics, date expressions, and so on, they can be understood as the results of semantic and pragmatic analysis specialized for statistical and/or event information. The specification of the annotation maintains the previous one, which was explained in [5]. In addition to the existing corpus on 582 articles on 27 topics, 120 articles on eight topics, used as the subjects of the T2N subtask, were newly annotated this time.

4.2 Visualization Platform

The visualization platform aims to make it easier and quicker to construct several types of information visualization systems and multi-modal presentation systems. It provides commonly used functions for developing visualization systems as libraries, such as drawing statistical charts, selecting a region to be depicted, and zooming/panning those. The visualization platform is expected not only to make it easier to develop a visualization system, but also to compare different visualization systems by giving them a standardized look-and-feel.

Figure 1 shows a structure of visualization systems constructed using the visualization platform. A developer who wants to build a new system is required only to design a visual structure, which is a logical configuration of visual components such as a chart and text box and functions that manipulate those, and to implement a visual structure constructor, which constructs the designed visual structure by combining contents to be visualized, such as statistical data and related texts extracted from corpuses. Then the visualization platform translates it into a view, which is the physical appearance of a given visual structure with a set of functions implemented. The current platform supports only statistical charts and Japanese map-based visualization as main visual components. The visual structure is represented in XML, which is interpreted by the platform, and it is possible to add a visual component unique to the system by embedding SVG (Scalable Vector Graphics) code into it. Such extension is allowed in order to strike a suitable balance between ease of building a visualization system and richness of generated visual components.

A generated view consists of several panes as shown in Figure 2, each of which corresponds to visual components specified in a visual structure. A graph pane, shown as Figure 2-(1), is used to display a statistical chart and annotations placed on the chart. The graph pane has a layered structure. Among the layers, a base layer and a control slider layer (corresponding to a zooming pane shown as Figure 2-(4)) are automatically appended by the platform. The other layers are created according to the visual structure specified, and the number of layers is not restricted. For instance, in Figure 2-(1), two layers for differect types of annotations, an estimated value layer and an annotation layer, are appended.

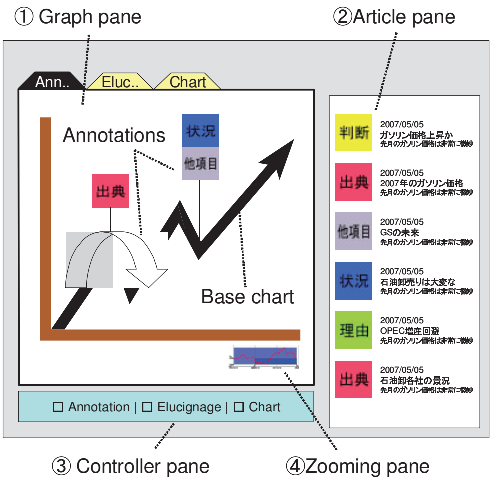


Figure 2. The Structure of Views

In a article pane, shown as Figure 2-(2), a list of articles related to the chart is displayed with snippets. These snippets are indicative summaries of those articles, and are used to help users decide at a glance whether or not an article is worth reading.

A control pane, shown as Figure 2-(3), is used to switch a layer to visible or non-visible. This pane is also automatically generated by the visualization platform.

Figure 3 shows snap-shots of trend information

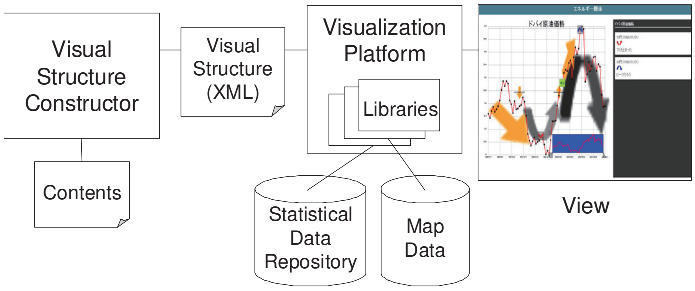


Figure 1. The Structure of Visualization System

generated by two visualization systems, STEND and Elucignage. STEND is a visualization system that extracts numeric information and qualitative information from a set of newspaper articles and visualizes them in a special chart. In contrast, Elucignage is a visualization system that shows numeric information along with icons, which represent related text information. These systems reflect the different interests on their visualization, while handling the same topic. In Figure 3-(1), these appearances are quite different and so it seems difficult to compare them. In contrast, Figure 3-(2) shows the views constructed using the visualization platform: the appearances of the two systems are similar and therefore they are much easier to compare.

4.3 Change Expression Corpus

The change expression corpus consists of expressions of changes extracted from newspaper articles and their formal representations, which can be referred to when designing and evaluating the extraction and visualization of statistical information from those texts. The principle of designing this corpus is that the basic components of extraction and visualization of statistical information from texts are descriptions of changes rather than values of each time point.

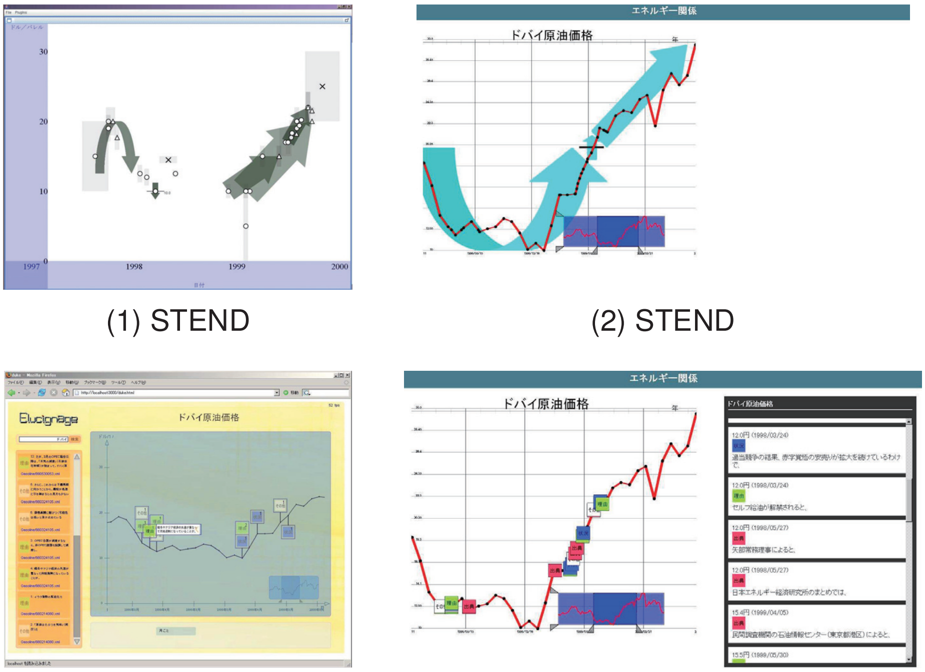
The following two text snippets are taken from newspaper articles on oil prices, in which direct translations of Japanese text into English are used for explanation.

> The oil price (Dubai Oil) has kept dropping since its peak last October, of around $20 per barrel, and fell to $12.50 in late January. It then rose temporarily, because of the tense situation in Iraq, but has been languishing recently at around $12, due to oversupply. (14/Feb/1989)
> The oil price, with Dubai Oil used as the index, surged to $18.30 on the 15th. This rise represents an increase of 80% from $10 in February. (17/Jul/1999)

These texts contain many qualitative descriptions of changes during some time periods and at some time points, such as “kept dropping” and “its peak.” Such descriptions are observed as frequently as those of statistical values at some time points, which constitute data points plotted on a chart. In a sense, descriptions of changes can be core information, and data points are subordinately shown as the start points and end points of changes. Even “primitive” information on data points is stated as a change of becoming that value at that point. In addition, these pieces of information on changes are more important than one of the data points for grasping an outline and trend of given time-series information. In linguistic realizations, also, information is articulated based on changes, such as “has kept dropping,” “since its peak,” “fell,” “rose,” and “has been languishing.” Representation using changes as the primitive unit is pursued based on these observations.

A piece of articulated information is usually stated as a text clause, and its main verb specifies the type of change and the relation to the value expressed as its argument. For example, “rose to $10” states that there has been an upward change and that the value of the statistic is equal to its argument $10 at its event time. Many verbs describing changes are characterized in these two axes: the type of change and the relation to the value. As for the types of changes, “was $10” states nothing about its change, “became $10” states that some change happened in an unspecified direction, and “stayed at $10” states that there was no change. As for the relations to the values, not only equality at the event time, but also many relations can be expressed by using various verbs. An upward change and an existence of a time point before the event time at which the value was equal to $10 is stated by “exceeded $10”, which implies that the value at the event time is above $10.

Verb groups and some verbs with a complicated meaning can state features on the third axis: change of changes, such as the start and end of change. The start of an upward change is represented by “started to increase,” and a change of direction by “rebound.” A similar type of statement on change of changes can be stated by “peak”, though it is not a verb. Table 2 shows typical expressions stating changes that are



(1) STEND (2) STEND

(1) Elucignage (2) Elucignage

Figure 3. Examples of Visualization

summarized and characterized by features organized in the three axes.

Some verbs, such as "surge" compared to "increase" and "languish" compared to "stay", have some implication or associated meaning. Using those words, qualitative evaluations are added to the basic information of changes organized in Table 2. In other words, the impacts and evaluations of the changes expressed can be recognized by paying attention to the usage of those verbs.

Based on this principle, the basic unit of information was designed to consist of the following constituents:

- The name of a statistic concerned
- The event time, which specifies the time (date) this unit of information is related/tied to
- Type 0 information, which describes the value of the statistics at the specified time point, which is not always the same at the event time
- Type 1 information, which describes the change in value of the statistics
- Type 2 information, which describes the change or rate of change

The names of information types are based on their relationship to the order of derivation.

Type 0 information is represented as a triplet consisting of the relation to the event time, the relation to the value, and the value to be related. The relation to the event time can be either "during," "before," or "after," which indicates the time when the relation to the value mentioned holds is during the event time, or preceding it, or succeeding it, respectively. When this relation is "during," the relation to the value can be either "equal," "larger," or "smaller," which respectively mean the value of the statistic is equal to, larger than, or smaller than the value mentioned. It is always "equal" when the relation to the event time is "before" or "after."

Type 1 information, which is about the change of values, can be "some-change" or "no-change" depending on whether there is some or no change in that statistic at that time. Upward and downward describe the situation of the change more specifically, when some change is happening, i.e. those are sub types of "some-change." When the type 1 information is "upward" or "downward," two parameters show the characteristics of that change: *diff* and *ref/dur*. diff shows the proportion or difference of the values made by that change, and ref/dur shows when that change started to take place or how long the change took.

Type 2 information, which is about the change or degree of changes, can be one of the following six options.

"continue" when the same type of change continues as before at that point

"start" when a new type of change, namely a change of a different type from the previous one, has

Table 2. Verbs and information on changes

Verbs	Relationship to the value/time	Type of change	Type of change of changes
was $10	equal/during		
reached $10	equal/during	some change	
stayed at $10	equal/during	no change	
increased (to $10)	(equal/during)	upward	
dropped (to $10)	(equal/during)	downward	
went over $10	equal/before	upward	
fell below $10	equal/before	downward	
approached $10	equal/after	some change	
was over $10	above/during		
was under $10	below/during		
started to increase		upward	start
started to decrease		downward	start
kept to increase		upward	continue
kept to decrease		downward	continue
rebounded		upward	change of direction
rallied to $10	equal/during	upward	change of direction

started at that point

"reverse" when a change of a direction opposite to that of the previous one has started at that point. This is a sub type of "start."

"convex" when there happened changes of opposite direction previously, that is, the change of value of the statistic shows a convex or concave shape, and in the past, there was a time point when the value was the same as the one of that point

"speed-up" when the degree of some change is getting stronger or accelerating at that point

"speed-down" when the degree of some change is getting weaker or slowing down at that point

The value of type 2 information has a parameter, *dur/ref*, which indicates the period of continuation of that change or length of the interval before the change re-started.

The change expression corpus has been constructed according to this coding system. 886 units of information were extracted and coded from 99 articles on five topics chosen from the existing MuST data set, and 912 units were extracted and coded from 120 articles on eight topics of the subjects of the T2N subtask. In total, the constructed change expression corpus contains 1789 units of information from 219 articles.

5 The Free Subtasks

The researches conducted as free subtasks can be divided into four categories. The researches in the first category, which have been motivated by the same interest to that of the N2T task, are working on generating from numerical data the text explaining it. Two teams are tackling this problem. Taking the trend of stock price as the theme, Kobayashi et al.[8] constructed a corpus of expressions appearing in that domain, and using this as the domain knowledge, selected and combined suitable verbal expressions according to the categorization of the shapes of changes to be expressed. Umano et al.[15], on the other hand, explained the changes by assigning verbal expressions that are domain independent and general, focusing on both global characteristics such as increasing or decreasing and local characteristics such as the degree of distribution. Combining these two types of knowledge and expressions, which are domain specific and domain independent, is the next step in order to make these methods widely applicable and to generate more fluent expressions.

The second category mainly concerns information visualization. Aiming at viewing earthquake information, which has both temporal and geographical characteristics, from several view points, Takama et al.[13] proposed a model named the visualization cube, which was obtained by applying a model used in OLAP to visualization. Ito et al.[2] made it possible to access information exploratively by clustering documents and keywords separately and arranging those clusters in two mutually related spaces. These researches are two examples of the interactive and explorative information access methods that the MuST workshop is pursuing. Iwata et al.[3] proposed an interesting method of representing geographical information. In addition to the difficult problem of representing multi-dimensional information, they are addressing the problem of representing vagueness of information, which is related to presentation of summaries and overviews.

The third category consists of researches on text mining and the visualization of information found through it. Kawai et al.[7] visualized the relations of statistical terms by calculating their co-occurrence frequencies, which they think characterize events and phenomena in the real world. The global dynamic network obtained this way allows us to overview the structure of complex and international problems. We need to know what structure a given problem has and what factors are related to it in order to access correct information about it. This research solves this problem. Murata et al.[10] allowed several relations described in the set of newspaper articles to be overviewed by extracting and plotting relationships between numerical values, numerical values and proper names, and also numerical values and nouns of a common semantic category. This proposal derives a novel method of overviewing all information contained in a given set of documents. Yoshida et al.[16] and Rzepka et al.[12] focused on stock prices and sentencing process, and tried to identify what factors affected those by analyzing texts. A way of integrating these types of text mining and information retrieval based on a more specific interest should be tackled in the next step.

The fourth category is related to the T2N subtask and consists of research on component technologies that can be used for it. Especially, identifying and extracting names of statistics are addressed by many groups, although different names are used such as attribute names and statistical terms. The works of Yoshida et al.[16] and Mori et al.[9] are examples of this type of research, as well as the above-mentioned work of Kawai et al.[7]

Every research reported as a free subtask is working on an important theme concerning the explorative process from overviewing an abstract to accessing specific information. In the future, it is important to construct a whole picture of this process by integrating those component technologies.

6 T2N Subtask

The T2N subtask requires systems to extract time-series statistical information from a set of texts, and to compile it into a given format. This makes it possible to identify what time point and what value of a given statistic are discussed in those texts. The data extracted can be used also for visualizing the statistical changes according to the interests expressed in those texts. In this sense, it is one of the important elements for compiling trend information. The T2N subtask of MuST in NTCIR-7 was designed to be a task of extracting a set of pairs consisting of a time point and the value of a given statistic at that time point, though there are various possibilities of the range of data to be extracted, and all types of information on changes coded in the change expression corpus could be useful when employing some sophisticated visualization method. Statistics in the subjects are identified by their name and unit expression. For some statistics, more than one name and/or unit expression is given, and these are grouped by related topics. Specifically, topics for subjects were decided first, then two to four statistics were chosen from each topic. The same set of articles was specified to be processed in each topic, and the articles were taken from the Mainichi newspaper from 1998 to 2001. Table 3 shows the subjects of the T2N subtask in this time. As for the former four topics, articles on the same topics were included in the existing MuST data set and also in the change expression corpus delivered before the formal run. Articles used for the formal run, however, were new. The latter four topics and articles handled were brand new to the participants. The period between the subject delivery to the participants and the submission deadline was around two weeks, and system tuning after the subject delivery was allowed.

Nine submissions from five teams were made for the formal run. The first evaluation was conducted using precision, recall and F-measure similar to ordinary information extraction tasks. Table 4 and Table 5 show micro averages of those grouped by each topic due to the space limitation. The effective numbers of extracted data show the number of data extracted excluding duplications and data violating the task specifications. Since this task requires systems to extract the values of specified statistics and associate them with the relevant time points, the errors are divided into two groups according to this process: errors that a system extracted wrong values, that is, the values of other statistics, and errors that a system wrongly associated the time point with the value extracted correctly. In the table, the numbers of *value errors* and *time errors* show the numbers of the former and latter types of errors respectively. Figure 4 depicts the F-measures in Table 4 and 5. Many systems obtained poor results in the topicof Industry; this was probably because the statistics handled in that topic are indexes and have no unit expression, and are thus a type of statistics that many systems do not attempt to handle.

Each system employs a different method that characterizes it. The Keio1 system[1], which achieved the best result overall, actively uses several kinds of information specific to a given statistic. Those include lists of words that should be included in a sentence describing the value of a given statistic and of words that should not. It also processes all texts first and estimates the range of the values that a given statistic could take, and then filters out values lying outside of that range. Values are associated with relevant time points by a heuristic search using dependency relations

Table 3. Subjects of the T2N Subtask

Topic	#articles	Statisitcs
Gasoline	12	pomp price of regular gasoline (national average)
		Dubai oil price
Personal Computer	10	value of national shipments
		volume of national shipments
Communication Device	19	number of cellular phone subscribers
		number of PHS subscribers
		total number of PHS and cellular phone subscribers
		number of land phone subscribers
Political Trend	16	approval rating for the Cabinet
		disapproval rating against the Cabinet
		approval rating for the LDP
		approval rating for the DPJ
Industry	17	industrial production index
		industrial shipping index
		industrial inventory index
I-mode	20	number of i-mode subscribers
		number of EZweb subscribers
		number of J-Sky subscribers
		number of cellular phone-based Internet connection
DigitalCamera	8	volume of national shipments
		value of national shipments
CenterExam	18	number of applicants
		ratio of successful applicants
		ratio of taking exams
		ratio of usage in private universities

among bunsetsus, which are units of Japanese sentences each consisting of one content word followed by a few functional words.

The Keio2 system[1], which showed good performance in many topics, uses a set of hand-crafted rules for categorizing bunsetsus and deciding the type of information in those, extracts elemental pieces of information, then organizes those using dependency relations among the bunsetsus. The number of rules exceeds 100, so these elaborate rules contribute to the good performance of the system.

The Forst system[9] extracts statistical values and time expressions using a machine learning method, by regarding those problems as a sequence labeling problem. Those elements are paired by referring to the distances counted by the number of characters between them. Then, those with the relevant unit expression are selected as the data of a given statistic. Because only unit expressions are used and not the names of statistics, the performance drops for topics that contain many statistics with the same unit expression. The method of association with time points is simple, too. Nevertheless, this system attains good performance in a few topics, probably because of its good ability to extract basic elements.

In the four systems submitted by the HCU team[11], the performances were examined for combinations of two methods proposed for extracting the values of a given statistic and two methods of assignment of time points. One method of time-point assignment is so simple that all time points of data are set to the date when the article handled was published. The fact that such a simple method did not suffer as large a drop in performance as more sophisticated methods reveals the difficulty of the problem handled in the T2N subtask. The two methods of extracting statistical values differ in the degree of the respect for precision or recall.

Unlike the others, the Mie system[14] tries to extract data implicitly expressed in texts by focusing on relative expressions. Relative expressions are identified using hand-crafted patterns. This system uses similarity of words obtained from a hierarchical thesaurus for judging to which statistic the value extracted belongs.

The approach taken by the deNdai system[16] is unique: it employs a method of extracting statistical names developed through their free subtask. First, all triplets of statistical name, statistical value, and time point, are collected from a given set of texts. Then, those triplets are divided into groups based on similarities of statistical names and statistical values. Using the similarity of statistical values is supported by the intuition that the values of the same statistic are not

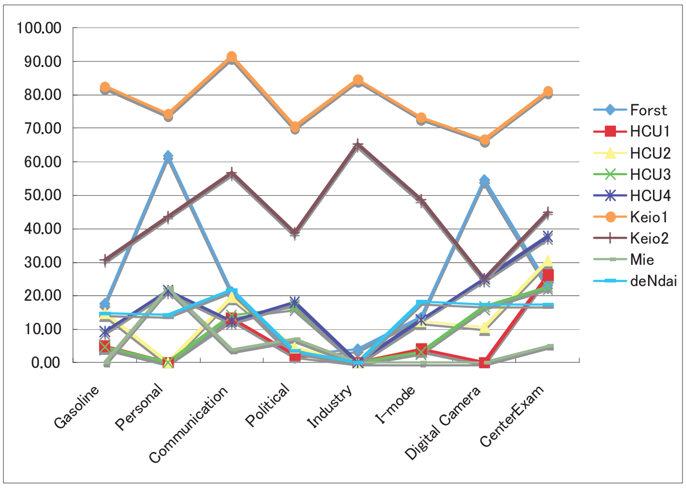


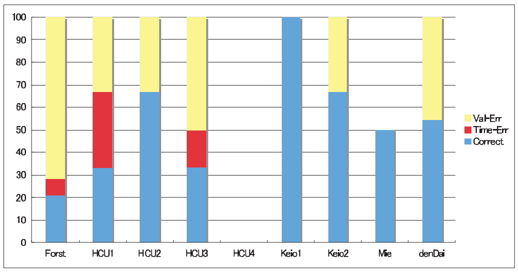
Figure 4. F-measures of the Submissions

so different and stay in some range. At this point, the system obtains groups of triplets, each of which corresponds to data on one statistic. This output is similar to that of the system of Murata et al.[10] Lastly, the system selects one group as the result on a given statistic using similarity of a given name and the names of statistics of triplets in the group. It is interesting that specific information is extracted by considering the whole set of information in texts. It reflects a characteristic of summarizing trend information, in which texts ranging over a considerable time span are considered as a whole.

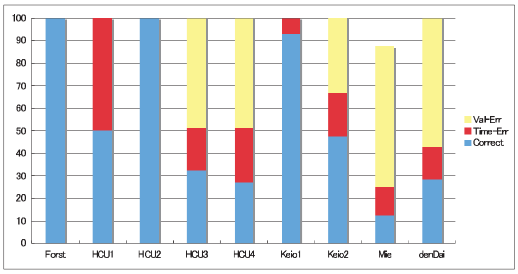
The T2N task differs from ordinary information extraction tasks, in which it is required to extract data of the same statistic over a considerable time range. This characteristic leads to possibility of employing special evaluation measures suitable to the task. The rest of this section discusses those possible measures, though we could use none of the measures this time.

Since the extraction is usually done in the two stages of extraction of values and assignment of time point, the errors can be divided into two categories based on which stage caused the error. This categorization could add new information to the measures related to precision. Figure 5 shows examples of such a three-valued evaluation. For example, in some topic, in which many statistics with the same unit expression are discussed, extraction of values could be difficult. This type of figure may be able to reveal such a characteristic of problems.

As for a measure related to recall, a chart could be used, which is drawn using data extracted correctly by a given system. Figure 6 shows three such charts on cabinet approval rating. Compared to chart (1), which is drawn using correct data extracted by humans, we can understand intuitively how precise systems, whose outputs are drawn as chart (2) and chart (3), reproduce the data. When a quantitative evaluation is needed,



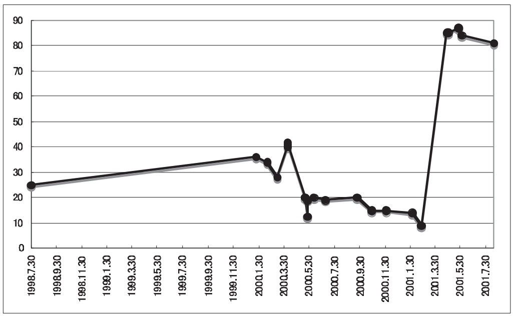
Number of cellular phone subscribers



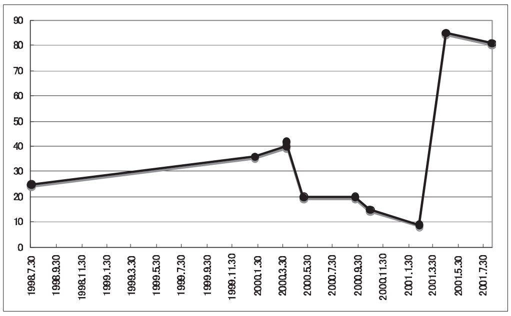
Approval rating for the Cabnet

Figure 5. Three Valued Precision

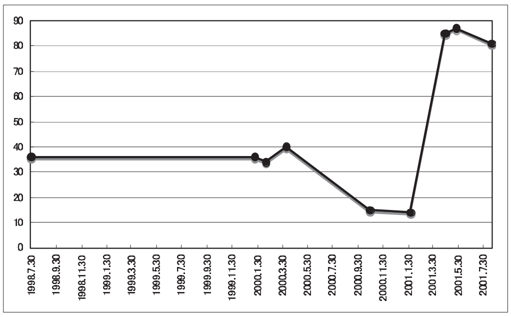
the mean square error to the correct chart can be used. There are two problems with this measure, however. First, since the range of values and unit are different for each statistic, such mean square errors cannot be compared between different statistics. Second, for a statistic whose values change linearly such as the number of cellular phone subscribers shown in Figure 7, such a chart can be reproduced from a few data points, and the recall rate of a system cannot affect the shape of the chart correctly. More consideration is needed to elaborate this idea into a viable general measure.



(1) Perfect Result



(2) System Output



(3) System Output

Figure 6. Chart-based Recall

7 Discussion

The MuST workshop has evolved in NTCIR-7. It has taken a first step as an evaluation workshop, though the number of subtasks was fewer than initially planned. The T2N subtask concerns a theme that has been tackled by many participants in the previous cycles. By providing common subjects and common evaluations, however, the characteristics of each system have become clearer. The next step is to examine the relationship between the characteristics of given subjects and the performance of systems. Some sophisticated evaluation measure for revealing that relationship should also be considered.

Meanwhile, researches on the free subtasks have also progressed steadily. Since it is difficult to evaluate researches quantitatively in some fields such as visualization and text mining, we should keep this structure of the workshop and make our discussions in the community more fruitful.

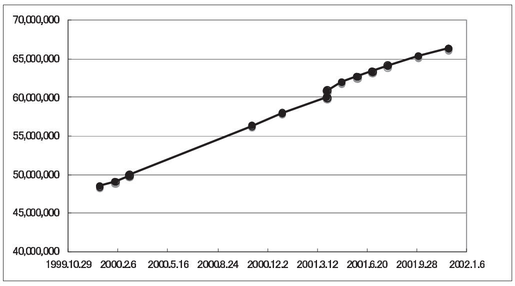


Figure 7. Cellar Phone Subscribers

Resources have been constructed, too. In addition to constructing new data sets, it is important to propose the visualization platform. Although we could not make it usable for the participants this time, it is expected to play an important role in researches in the field.

8 Conclusion

The overview of MuST at NTCIR-7 was reported in this paper. The workshop has two aspects: it is an evaluation workshop, and also a community with a shared interest. In the former role, it assesses the state of common component technology. In the latter role, it showcases mutually related researches in the field. Resources have been constructed steadily. These activities are assisting researches on multi-modal summarization of trend information.

References

[1] Tatsukuni Inoue, Takashi Yamamoto, Makoto Toriyabe, and et al. Extraction of Chronological Statistics Using Domain Specific Knowledge, in this proceedings.

[2] Takayuki Itoh and Haruho Tachibana. Visualization of Corpus Data by a Dual Hierarchical Data Visualization Technique, in this proceedings.

[3] Kenichi Iwata and Mariko Sasakura. A Method to Visualize NUmerical Data with Geographical Information using Feathered Circles Painted by Color Gradation, in this proceedings.

[4] Tsuneaki Kato, Mitsunori Matsushita, and Noriko Kando. MuST: A Workshop on Multimodal Summarization for Trend Information, Procs of the Fifth NTCIR Workshop Meeting on Evaluation of Information Access Technologies, pp. 556 - 563 (2005).

[5] Tsuneaki Kato, Mitsunori Matsushita, and Noriko Kando. Expansion of Multimodal Summarization

for Trend Information –Report on the First and Second Cycles of the MuST Workshop –. Procs. of the Sixth NTCIR Workshop Meeting, pp. 235–242 (2007).

[6] Tsuneaki Kato, Mitsunori Matsushita, and Noriko Kando. Fostering Multi-Modal Summarization for Trend Information. Procs. of 11th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2007), Part 2, pp. 377–386 (2007).

[7] Hideki Kawai, Haruka Saito, Masaaki Tsuchida, and et al., Visualization for Statistical Term Network in Newspaper, in this proceedings.

[8] Ichiro Kobayashi and Naoko Okumura. Text Generation for Explaining the Behavior of 2D Charts: With an Example of Stock Price Trends, in this proceedings.

[9] Tatsunori Mori and Rintaro Miyazaki. A simple baseline method for NTCIR-7 MuST T2N task: Yokohama National University at NTCIR-7 MuST T2N, in this proceedings.

[10] Masaki Murata, Tamotsu Shirado, Kentaro Torisawa, and et al. Sophisticated Text Mining System for Extracting and Visualizing Numerical and Named Entity Information from a Large Number of Documents, in this proceedings.

[11] Hidetsugu Nanba. Extraction of Trend Information from Newspaper Articles: Hiroshima City University at NTCIR-7 MuST, in this proceedings.

[12] Rafal Rzepka, Masafumi Matsuhara, Yasutomo Kimura, and et al. Toward Automatic Support For Japanese Lay Judge System - Processing Precedent Factors For Sentencing Trends Discovery, in this proceedings.

[13] Yasufumi Takama and Takashi Yamada. Interactive Information Visualization of Trend Information, in this proceedings.

[14] Yasuhiro Uenishi, Tatsuaki Matsuba, Fumito Masui, and et al. Trend Information Extraction based on Relative Expression participated on MuST T2N Task, in this proceedings.

[15] Motohide Umano and Naoyuki Koizumi. Verbal Expression of Time Series with Global Trend and Local Features, in this proceedings.

[16] Minoru Yoshida, Takahiro Sugiura, Takamasa Hirokawa, and et al. TDU Systems for MuST: Attribute Name Extraction, Text-Based Stock Price Analysis, and Automatic Graph Generation, in this proceedings.

Table 4. Evaluation of Submissions of the T2N Subtask

Run	Topic	#answer	#extrct	#effective	#crrct	time-err	val-err	Prec	Recall	F
	Gasoline	33	36	35	6	15	14	17.14	18.18	17.65
	PC	19	37	36	17	4	15	47.22	89.47	61.82
	CommDevice	50	211	188	25	9	154	13.30	50.00	21.01
Forst	Political	85	4	4	1	0	3	25.00	1.18	2.25
	Industry	27	33	23	1	0	22	4.35	3.70	4.00
	I-mode	45	305	220	18	18	184	8.18	40.00	13.58
	DigitalCamera	18	26	26	12	2	12	46.15	66.67	54.55
	CenterExam	37	41	41	9	3	29	21.95	24.32	23.08
	Gasoline	33	8	8	1	3	4	12.50	3.03	4.88
	PC	19	0	0	0	0	0	0.00	0.00	0.00
	CommDevice	50	11	11	4	3	4	36.36	8.00	13.11
HCU1	Political	85	2	2	1	1	0	50.00	1.18	2.30
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	4	4	1	2	1	25.00	2.22	4.08
	DigitalCamera	18	1	1	0	1	0	0.00	0.00	0.00
	CenterExam	37	9	9	6	2	1	66.67	16.22	26.09
	Gasoline	33	8	8	3	1	4	37.50	9.09	14.63
	PC	19	0	0	0	0	0	0.00	0.00	0.00
	CommDevice	50	11	11	6	1	4	54.55	12.00	19.67
HCU2	Political	85	2	2	2	0	0	100.00	2.35	4.60
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	4	4	3	0	1	75.00	6.67	12.24
	DigitalCamera	18	1	1	1	0	0	100.00	5.56	10.53
	CenterExam	37	9	9	7	1	1	77.78	18.92	30.43
	Gasoline	33	10	10	1	3	6	10.00	3.03	4.65
	PC	19	3	3	0	0	3	0.00	0.00	0.00
	CommDevice	50	21	21	5	3	13	23.81	10.00	14.08
HCU3	Political	85	186	181	22	18	141	12.15	25.88	16.54
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	17	17	1	5	11	5.88	2.22	3.23
	DigitalCamera	18	6	6	2	2	2	33.33	11.11	16.67
	CenterExam	37	16	16	6	5	5	37.50	16.22	22.64
	Gasoline	33	10	10	2	2	6	20.00	6.06	9.30
	PC	19	9	9	3	0	6	33.33	15.79	21.43
	CommDevice	50	15	15	4	1	10	26.67	8.00	12.31
HCU4	Political	85	186	181	24	16	141	13.26	28.24	18.05
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	17	17	4	2	11	23.53	8.89	12.90
	DigitalCamera	18	6	6	3	1	2	50.00	16.67	25.00
	CenterExam	37	16	16	10	1	5	62.50	27.03	37.74

Table 5. Evaluation of Submissions of the T2N Subtask

Run	Topic	#answer	#extrct	#effective	#crrct	time-err	val-err	Prec	Recall	F
Keio1	Gasoline	33	30	30	26	4	0	86.67	78.79	82.54
	PC	19	16	16	13	0	3	81.25	68.42	74.29
	CommDevice	50	45	44	43	0	1	97.73	86.00	91.49
	Political	85	51	51	48	1	2	94.12	56.47	70.59
	Industry	27	25	25	22	3	0	88.00	81.48	84.62
	I-mode	45	26	26	26	0	0	100.00	57.78	73.24
	DigitalCamera	18	15	15	11	4	0	73.33	61.11	66.67
	CenterExam	37	46	32	28	4	0	87.50	75.68	81.16
Keio2	Gasoline	33	19	19	8	8	3	42.11	24.24	30.77
	PC	19	13	13	7	3	3	53.85	36.84	43.75
	CommDevice	50	33	31	23	0	8	74.19	46.00	56.79
	Political	85	122	121	40	11	70	33.06	47.06	38.83
	Industry	27	25	22	16	3	3	72.73	59.26	65.31
	I-mode	45	39	25	17	2	6	68.00	37.78	48.57
	DigitalCamera	18	6	6	3	3	0	50.00	16.67	25.00
	CenterExam	37	34	34	16	9	9	47.06	43.24	45.07
Mie	Gasoline	33	0	0	0	0	0	0.00	0.00	0.00
	PC	19	8	8	3	0	2	37.50	15.79	22.22
	CommDevice	50	2	2	1	0	0	50.00	2.00	3.85
	Political	85	28	28	4	7	16	14.29	4.71	7.08
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	0	0	0	0	0	0.00	0.00	0.00
	DigitalCamera	18	0	0	0	0	0	0.00	0.00	0.00
	CenterExam	37	3	3	1	0	2	33.33	2.70	5.00
deNdai	Gasoline	33	9	8	3	3	2	37.50	9.09	14.63
	PC	19	9	9	2	5	2	22.22	10.53	14.29
	CommDevice	50	33	33	9	1	23	27.27	18.00	21.69
	Political	85	34	30	2	2	26	6.67	2.35	3.48
	Industry	27	0	0	0	0	0	0.00	0.00	0.00
	I-mode	45	25	21	6	2	13	28.57	13.33	18.18
	DigitalCamera	18	5	5	2	1	2	40.00	11.11	17.39
	CenterExam	37	9	9	4	1	4	44.44	10.81	17.39