A Summarization System for Chinese News from Multiple Sources

Hsin-Hsi Chen, June-Jei Kuo, Sheng-Jie Huang, Chuan-Jie Lin, and Hung-Chia Wung Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, ROC. E-mail: hh_chen@csie.ntu.edu.tw

This article proposes a summarization system for multiple documents. It employs not only named entities and other signatures to cluster news from different sources, but also employs punctuation marks, linking elements, and topic chains to identify the meaningful units (MUs). Using nouns and verbs to identify the similar MUs, focusing and browsing models are applied to represent the summarization results. To reduce information loss during summarization, informative words in a document are introduced. For the evaluation, a guestion answering system (QA system) is proposed to substitute the human assessors. In large-scale experiments containing 140 questions to 17,877 documents, the results show that those models using informative words outperform pure heuristic voting-only strategy by news reporters. This model can be easily further applied to summarize multilingual news from multiple sources.

Introduction

Owing to the widespread use of the Internet, a large scale of multicultural and multilingual information can be obtained simultaneously by computer users despite geographic differences. At such an information explosion age, how to filter useless information, and how to absorb and employ information effectively becomes an important issue for users. Take news broadcasting as an example. On-line services become popular. Readers can access the on-line news sites quickly, but news take much time for people to read. This article will provide a personal news secretariat to help on-line readers to absorb news information from multiple sources. This news secretariat eliminates the redundant information in the news, and reorganizes the relevant news for readers.

Reorganizing news is a kind of documents' summarization, which extracts important information for readers. Besides this application, document summarization can also help users to decide relevance of documents. This will

Received July 2, 2002; revised January 31, 2003; accepted April 14, 2003

eliminate some degree of bottlenecks on the information highway. The research of document summarization begins very early (Edmundson, 1964, 1969; Luhn, 1958), and becomes one of the traditional topics in the research of natural language processing. Recently, it has attracted new attention due to the Internet application. Many papers about documentation summarization have been proposed (Brunn, Chali, & Pinchak, 2001; Hovy & Marcu, 1998; Kupiec, Pedersen, & Chen, 1995; Lin & Hovy, 1997). A special summarization evaluation Summac-1 (Mani et al., 1998) organized by DARPA Tipster Text Program was held to deal with three kinds of evaluation tasks, i.e., categorization, ad hoc, and question and answering. In the past, the major research was stressed upon single document summarization. Recently, the efforts transferred to multiple documents summarization (Carbonell & Goldstein, 1998; Chen & Huang, 1999; Lin & Hovy, 2002; Mani & Bloedorn, 1997; McKeown, Klavans, Hatzivassiloglou, Barzilay, & Eskin, 1999; McKeown & Radev, 1995; Radev & McKeown, 1998; Radev, Blair-Goldensohn & Zhang, 2001).

Our research on single document summarization in Summac-1 (K.H. Chen et al., 1998) is extended to multiple documents summarization in Chinese news. This article is organized as follows: The next section presents the architecture of our summarization system. Then we specify a news clusterer. After that we deal with a news summarizer with informative words and its experimental results. Similarity analysis and presentation models are discussed, respectively. We then present a Question Answering (QA) system and introduce a new automatic evaluation model. Manual evaluation and automatic evaluation are compared. The next section then shows a large-scale experiment using a QA system, in that both two metrics, i.e., document retention rate, and QA correct rate, are considered. Then we deal with some issues on multilingual news summarizer. Finally, we show the concluding remarks.

Architecture of a Summarization System

Figure 1 shows the architecture of our summarization system, which is used to summarize Chinese news from

^{© 2003} Wiley Periodicals, Inc.



FIG. 1. Architecture of a summarization system.

on-line newspapers, for example, *China Times, Commercial Times, China Times Express, United Daily News, Tomorrow Times,* and *China Daily News* in Taiwan. It consists of two major components: a news clusterer and a news summarizer. The news clusterer receives news streams from multiple on-line news sites and directs them into several output news streams according to events. An event is denoted by five basic entities such as people, affairs, time, places, and things. The news articles for respective events are summarized by a news summarizer. The tasks for the clusterer are listed below:

- 1. Employing a segmentation system to identify Chinese words.
- 2. Extracting named entities, for example, people, place, organization, time, date, and monetary expressions.
- 3. Applying a tagger to determine the part of speech for each word.
- 4. Clustering the news stream based on the named entities and other signatures.

The tasks for the news summarizer are shown as follows:

- 1. Partitioning a Chinese text into several meaningful units (MUs), which will be described later.
- 2. Linking the meaningful units and denoting the same thing from different news reports.
- 3. Displaying the summarization results by two kinds of modes: a sequence of news by information decay and a summarization by voting from reporters, which will be described later.

A News Clusterer

Because a Chinese sentence is composed of characters (words) without boundaries, segmentation is indispensable. We employ a dictionary, some morphological rules, and an ambiguity resolution mechanism for the segmentation. Furthermore, we also extract named organizations, people, and locations, along with date/time expressions and monetary and percentage expressions (H.H. Chen et al., 1998). Different types of information from different levels of text are also considered, including characters condition, statistic information, titles, punctuation marks, organization and location keywords, speech-act and locative verbs, cache, and n-gram model. The recall rates and the precision rates for the extraction of person names, organization names, and location names in the Chinese named entity extraction task of a famous message understanding competition (MUC, 1998) are (87.33%, 82.33%), (76.67%, 79.33%), and (77.00%, 82.00%), respectively.

When we apply the segmentation system in the summarization experiments, several errors occurred as below, that could influence the performance of summarization:

- 1. Two sentences denoting the similar meaning may be segmented differently due to the segmentation strategies. For example:
 - (C1) 但法務部長城仲模內定升任司法院副 院長 (But the justice minister Mr. Chung-Mo Cheng has been internally appointed as the vice president of

the Judicial Yuan . . .) This sentence is segmented into 但 法務部(Nc)

長城(Nc) 仲模(Nb) 內定(VC) 升任(VG) 司法院(Nc) 副院長(Na)...

(C2) 而城仲模轉任司法院副院長之後的法務部長遺缺

(... and after Mr. Chung-Mo Cheng has been appointed as the vice president of the Judicial Yuan, the vacant position of Justice Minister)

This sentence is segmented into 而城仲模(Nb)轉 任(VG) 司法院(Nc) 副院長(Na) 之後(Ng) 的法務(Na) 部長(Na) 遺缺(Na)... The major title "法務部長" (Justice Minister) and the

major person "城仲模" (Chung-Mo Cheng) are segmented in different ways in the above examples. That will introduce errors in similarity analysis.

2. Unknown words generate many single-character words. For example, "上 (Na) 石(Na) 流(VC)," "園(Nc) 山(Na) 村(Nc)," "芭 (Nb) 比(VC) 絲(Na)," "老 (VH) 丙(Neu) 建(VC)," and so on. After tagging, these words tend to be nouns and verbs, which are used in computing the scores for similarity measure. Thus, errors will be illuminated later.

We adopt a two-level approach to cluster the news from multiple sources. At first, news is classified on the basis of a predefined topic set. Then, the news articles in the same topic set are partitioned into several clusters according to named entities. Clustering is necessary. On the one hand, a famous person may appear in many kinds of news stories. For example, President Bush may make a public speech (political news), join an international meeting (international news), or even just show up at the opening ceremony of a baseball game (sports news). On the other hand, a common name is frequently seen but denotes different persons. Clustering helps to reduce the ambiguity introduced by famous persons and/or common names.

A News Summarizer

Meaning Units

The basic idea in our study is to tell the similarity of the news articles in the same event. The basic unit for similarity checking may be a paragraph or a sentence. For the former, text segmentation is necessary for documents without paragraph markers (K.H. Chen & Chen, 1995). For the latter, text segmentation is necessary for languages such as Chinese. Because Chinese writers often assign punctuation marks at random (H.H. Chen, 1994), the sentence boundary is not clear. For Example:

> (C3)儘管警方大肆鎮壓與逮捕,無數反 對自由貿易的示威群眾今天仍繼續 向西雅圖市中心前進,他們發動和平 集會,以抗議世界各國貿易部長即將 在西雅圖舉行討論全球貿易自由化的 會議。

(Central News Agency, 1999.12.02) (Although they were undeterred by mass arrests and a police crackdown, antifree-trade protesters still marched on downtown Seattle today. The protesters, carrying signs and chanting, opposed the global trade liberalization being worked on at a meeting of trade ministers from the World Trade Organization.)

It is composed of four sentence segments separated by commas. On the one hand, if a sentence segment is regarded as a unit for similarity checking, it may contain too little information. On the other hand, if a sentence is regarded as a unit, it may contain too much information. To avoid such extreme results, we consider a meaningful unit (MU) as a basic unit for measurement. A MU is composed of several sentence segments and denotes a complete meaning. Following is the examples of our applying two MUs for (C3):

> (C4) 儘管警方大肆鎮壓與逮捕,無數反對 自由貿易的示威群眾今天仍繼續向西 雅圖市中心前進 (Although they were undeterred by mass arrests)

and a police crackdown, antifree-trade protesters still marched on downtown Seattle today.)

(C5) 他們發動和平集會,以抗議世界各國貿易部長即將在西雅圖舉行討論全球 貿易自由化的會議 (The protesters, carrying signs and chanting, opposed the global trade liberalization being worked on at a meeting of trade ministers from the World

Here, a MU that consists of several sentential segments denotes a complete meaning.

Trade Organization.)

Three kinds of linguistic knowledge—punctuation marks, linking elements and topic chains, are applied for identifying the MUs.

- Punctuation marks: There are fourteen marks in Mandarin Chinese (Yang, 1981), but only period, question mark, exclamation mark, comma, semicolon, and caesura mark are employed. The former three are sentence terminators, and the latter three are segment separators.
- 2. Linking elements: There are three kinds of linking elements (Li & Thompson, 1981): forward-linking elements, backward-linking elements, and couple-linking elements. A segment with a forward-linking (backward-linking) element is linked with its next (previous) segment. A couple-linking element is a pair of words that exists in two segments. Apparently, these two segments are joined together. In other words, these two segments linked by the above linking elements are obviously related in theme. Moreover, those linking elements are stored in a table form. Whenever the linking elements are identified by consulting the table during Chinese sentence processing, the related two sentence elements are extracted as one MU. Examples (C6)–(C8) show their respective linkings.
 - (C6) Forward linking:

因為天氣不好, 飛機改在明天起飛。 (Because the weather is bad, the flight is changed to tomorrow.)

- (C7) Backward linking: 我想早一點來,可是我沒趕上公車。 (I wanted to come earlier, *but* I didn't catch the bus.)
- (C8) Couple linking: 他<u>一邊</u>走路,<u>一邊</u>唱歌。 (He is *walking and singing*.)
- 3. Topic chains: The topic of a clausal segment is deleted under the identity with a topic in its preceding segment. The result of such a deleting process is a *topic chain*. In other words, a referent is referred to in the first clause, and then there follow several more clauses talking about the same referent but not overtly mentioning that referent. For this reason, we have the following postulation (Chen, 1994): given two VP segments, or one S and one VP segment, if their expected subjects are unifiable, then the two segments can be linked. This article applies the postulation to parse each segment independently, and composes a parsing tree from the trees of segments. This will reduce the complexity of parsing very long Chinese sentences. In our summarization system, we do not actually parse the Chinese sentences. We employ part of speech information only to predict a missing subject of a verb. If it does happen, we postulate that it must appear in the previous segment, at the same time the two segments are connected to form a larger unit. For example:
 - (C9) 國民黨是靠組織起家的政黨,現在的 組織體質卻很虛弱,所以選戰最後也 要仰仗文宣,實在很可惜。 (The Kuomintang (KMT) was built up by its orga-

nized system, but now the organized physique is very weak, so that in the end they had to depend on propaganda during the election campaign, which is really a pity.)

There are four segments in this example. The words ""卻"" (but) and ""所以"" (so that) in the second and the third segments are backward linking elements. Thus, these three segments are connected together. The last segment does not have any subject; therefore, it is connected to the previous one. To sum up, these four segments form a MU. Here, we do not touch on the associative problem mentioned in parsing (Chen, 1994). We only consider which segments can form a MU.

Similarity Models

The next step is to find the similarity among MUs in the news articles that report about the same event, and link the similar MUs together. Our approach uses linguistic features extracted from input documents to identify several groups of MUs which convey approximately the same information. Because predicate-argument structure forms the kernel of a sentence, verbs and nouns are selected for similarity measures. We do not employ machine learning algorithm to induce decision rules for text similarity (Hatzivassiloglo, Klavans, & Eskin, 1999). Instead, we consider several strategies shown as follows:

- (S1) Nouns in one MU are matched to nouns in another MU, as are the verbs.
- (S2) The operations in (1) are exact matches.
- (S3) A Chinese thesaurus, i.e., 同義詞詞林(tong2yi-4ci2ci2lin2), (Mei et al., 1982), is employed during the matching. That is, the operations in (S1) may be relaxed to inexact matches.
- (S4) Each term specified in (S1) is matched only once.
- (S5) The order of nouns and verbs in MU is not taken into consideration.
- (S6) The order of nouns and verbs in MU is critical, but it is relaxed within a window.
- (S7) When continuous terms are matched, an extra score is added to the similarity measure.
- (S8) When the objects of transitive verbs are not matched, a score is subtracted from the similarity measure.
- (S9) When date/time expressions and monetary and percentage expressions are matched, an extra score is added to the similarity measure.

The similarity measure of two MUs is in terms of nounsimilarity and verb-similarity shown as formulas (1) and (2), respectively. Besides Chen and Huang (1999), Hatzivassiloglou et al. (1999) also employed the same formulas.

noun
$$-\sin(A, B) = \frac{m}{\sqrt{ab}}$$
 (1)

$$\operatorname{verb} - \sin(A, B) = \frac{n}{\sqrt{cd}}$$
(2)

where m(n) denotes the number of matched nouns (verbs), a and b denote total number of nouns in MUs A and B, respectively, and c and d denote total number of verbs in MUs A and B, respectively.

Five models shown below are constructed under various combinations of the strategies specified above:

(M1) strategies (S1)+(S3)+(S4)+(S5) (M2) strategies (S1)+(S3)+(S4)+(S6)(M3) strategies (S1)+(S3)+(S4)+(S5)+(S7)+(S8)(M4) strategies (S1)+(S3)+(S4)+(S5)+(S7)+(S8)+(S9)(M5) strategies (S1)+(S2)+(S4)+(S5)+(S7)+(S8)+(S9)

Consider the following two MUs as an example for M1 model:

- MU1: 國父紀念館(Nc) 展出(VC) 的 國父(Na) 書畫(Na) 墨寶(Na), 傳出(VC) 了 失蹤(VH) 的 消息(Na) (The painting and calligraphy of Dr. Sun Yat-sen that were exhibited in the National Dr. Sun Yat-sen Memorial Hall have disappeared.)
- MU2: 國父紀念館(Nc) 傳出(VC) 了 國 **父**(Na) 書書(Na) 失蹤(VH) 的 消息(Na) (The National Dr. Sun Yat-sen Memorial Hall passed a message that the exhibiting painting and calligraphy have disappeared.)

In this example, m = 4, n = 2, a = 5, b = 4, c = 3, and d = 2. Thus, noun-sim(MU1,MU2) = 0.89, and verbsim(MU1,MU2) = 0.82. The basic ideas for the other models are dealt with in the following section.

Preparation of Testing Corpus

Nine events, which occurred within 1998/11/7 and 1998/ 12/8, were manually selected from Central Daily News, China Daily Newspaper, China Times Interactive, and FTV News Online in Taiwan for measuring the performance of each model. Each event was composed of more than two articles, which were reported in the same day. They are shown as follows:

- 1. 社會役的實施 (military service): six articles
- 2. 老丙建建築 (construction permit): four articles
- 3. 三芝鄉土石流 (landslide in Shan Jr): six articles
- 總統布希之子 (Bush's sons): four articles
 芭比絲颱風侵台 (Typhoon Babis): three articles
- 6. 股市穩定基金 (stabilization fund): five articles
- 7. 國父墨寶失竊案(theft of Dr. Sun Yat-sen's calligraphy): three articles
- 8. 央行調降利率 (interest rate of the Central Bank): three articles
- 9. 內閣總辭問題 (the resignation issue of the Cabinet): four articles

The news events are selected from different boards. An annotator reads all the news articles, and connects the MUs

TABLE 1. Performance of similarity of MUs.

Recall	
434	
905	
888	
198	
579	
5	

that discuss the same story. Because each MU is assigned a unique ID, the linking MUs form the answer keys for the performance evaluation.

Experiment Results

Traditional precision and recall are computed. Table 1 lists the performance of these five models. The thresholds for noun-similarity and verb-similarity are set to 0.3. M1 is regarded as a baseline model. M2 is different from M1 in that the matching order of nouns and verbs are kept conditionally. It tries to consider the subject-verb-object sequence. The experiment shows that the performance is bad. The major reason is the syntax of Chinese sentences is loose. We can express the meaning of a sentence by different syntactic structures. Movement transformation such as topicalization, relativization, ba-construction and bei-construction may affect the order of subject-verb-object. Thus, in M3 we give up the order criterion, but add an extra score when continuous terms are matched, and subtract some score when the object of a transitive verb is not matched. Compared with M1, the precision is a little higher, and the recall is improved about 4.5%. If we further consider some special named entities such as date/time expressions and monetary and percentage expressions in M4, the recall is increased about 7.6% at no expense of precision. M5 tries to estimate the function of the Chinese thesaurus. It uses exact matching. The precision is a little higher, but the recall is decreased about 6% compared with M4.

Several major errors affect the overall performance. Using nouns and verbs to find the similar MUs is not always workable. The same meaning may not be expressed in terms of the same words or synonymous words. Examples (C10) and (C11) talk about the same event, but use different verbs, such as 3 ± 7 (VC) "carry out" versus 2 [β](VC) "hold." The similarity contributed from verb is 0; thus, the recall is affected.

- (C10) 為了 宣示(VE) 穩定(VHC) 股市(Nc) 進行(VC) 第(Neu) 一(Neu)次 的 穩定 (VHC) 股市(Nc) 專案(Na) 小組(Na) 會議(Na) (In order to declare stability of stock market, the special subcommittee for stabling the stock market carried out the first meeting.)
- (C11) 專案(Na) 小組(Na) 昨日(Nd) 召開(VC)第(Neu) —(Neu) 次 會議(Na) (The special subcommittee held the first meeting yesterday.)

Besides, we can use different formats to express monetary and percentage expressions. For example, "二千八百三十億元" (two hundreds and eighty-three billion) in Chinese can be written as "二八三〇億元" or "2830 億." Similarly, the percentage expression "百分之七點二五" (7.25%) can be expressed in terms of "七•二五%" or "7.25%." Like other Asian languages, for example, Japanese, segmentation is another source of errors. Although our segmentation system takes care of named entities, there are still many new words in news articles. The new invented words are segmented into a sequence of single-character words. Moreover, the dictionary used in segmentation and the thesaurus used in the inexact matching are not integrated together in our experiments. A total fo 40% of nouns and 21% of verbs are not found in the thesaurus.

Presentation Models

How to minimize the redundacy between passages and how to select the most representative sentences are important issues in automatic summarization. Furthermore, sentence ordering in the summary is another important issue. There are many related approaches such as Maximun Marginal Relevance Muti-Document metric (Goldstein et al., 2000), exchange method (Hatzivassiloglou et al., 2001), Buddy system (Lin & Hovy, 2002), and sentence ordering (Barzilay, Elhadad, & McKeown, 2001). In this article, two models, i.e., focusing model and browsing model, are proposed to display the summarization results. In the focusing model, a summarization is presented by voting from reporters. For each event, a reporter records a news story from his own viewpoint. A news article is composed of several MUs. Those MUs that are similar in a specific event are common focuses of different reporters. In other words, they are worthy of reading. In the current implementation, the MUs that are reported more than twice are our target. For readability, the original sentences that cover the MUs are selected. For each set of similar MUs, only the longest sentence is displayed. The display order of the selected sentences is determined by relative position in the original news articles. That is, if a sentence appears in the introduction part of a news article, it tends to be displayed in the front of focusing summarization. The appendix lists the focusing summarization for typhoon news (event 5).

In the browsing model, the news articles are listed by information decay. The first news article is shown to the user in its whole content. The latter shown news articles, the MUs denoting the information mentioned before are shadowed (or eliminated), so that the reader can focus on the new information. The amount of information in a news article is measured by the number of MUs, so that the article that contains more MUs is displayed before the others. For readability, a sentence is a display unit. The appendix also demonstrates the browsing summarization. In this model, users can read both the reporters' common views and different views. It saves the reading time by listing the common view only once.

Discussion

Some issues of the above summarization system will be addressed as follows:

- 1. Goldstein et al. (1999) mentioned that summary length depends on the document type and fixed compression ratio is impractical. The summarization size of the above baseline system is fixed and cannot be used to study the variance between the length and the precision rate on Chinese newswire documents.
- The presentation order of sentences in a summary is based on the relative positions in the original documents despite of their importance. Thus, users might stop reading or miss the deferred appearing information.
- 3. The voting strategy gives a shorter summarization, which missed unique information reported only once.

To tackle the above problems, informative words concept is adopted. The extraction of the related informative words and the sentence selection methodologies for summarization are described in the following section.

Generating Summaries with Informative Words

The concepts of topic words and event words were applied to topic tracking successfully (Fukumoto & Suzuki, 2000). The basic hypothesis is that an event word associated with a story appears across paragraphs, but a topic word does not. In contrast to the event word, the topic word frequently appears across all documents. Thus, the document frequency of each word becomes an important factor in finding the appropriate sentences for summarization. The event words, which appear with higher term frequency in a document, will be more distinctive for the document. Therefore, we defined the words that have both high document frequency and high term frequency as informative words. They are used to improve the performance of news summarizer of the above baseline summarization system.

Informative Words and Sentence Selection for Summarization

The score function (*IW*) of an informative word W_{id} is defined as (5). $Ntf(W_{id})$ is the normalized term frequency of the term W_{id} . $tf(W_{id})$ and mtf(d) are the term frequency of W_{id} , and the man term frequency in document *d*, respectively. $DF(W_{id})$ denotes the document frequency of W_{id} , and *N* is the total number of documents in an event. In formula (5), λ denotes a weighted number that may be learned from a corpus. λ was set to 1/2 and 1 in the later experiments.

$$Ntf(W_{id}) = \frac{tf(W_{id}) - mtf(d)}{tf(W_{id}) + mtf(d)}$$
(3)

$$NDF(W_{id}) = DF(W_{id})/N \tag{4}$$

$$IW(W_{id}) = \lambda^*(3) + (1 - \lambda)^*(4)$$
(5)



FIG. 2. Example of QA task.

The more informative words a MU contains, the more possible the MU is used for generating summaries. In this article, only the top 10 terms with the higher IW scores will be chosen as informative words for a document. The score of each MU denotes the total number of informative words in it. The MUs with the highest score will be selected. Moreover, the selected MUs in a summary will be arranged in the descending order. In other words, the sentences that have more important MUs will appear before the less important ones in a summary. In this case, if the readers stop reading the summaries half way, they would not miss too much important information.

Experiment Results

Figure 2 shows our block diagram of the extrinsic evaluation task (Tsutomo, Sasaki, & Isozaki, 2001) on text summarization which was also adopted in the SUMMAC Q&A evaluation (SUMMAC, 1998). For simplicity, we call it *QA task*. First, the question sets (query sets) are collected under the document collection. At the same time, the corresponding answer sets are made after reading all the documents. After various kinds of document summaries are completed, the assessors will be involved in the evaluation. Each assessor will be assigned for summary texts and their related question sets. In the process of the evaluation, the reading and the answering time will be recorded. When assessors finish the question and answering task, we review their answers based on the respective answer sets and compute the precision rate of each question. Besides, the average document reduction rate and the average Q&A precision of various types of summary text are computed, respectively.

In our experiment, the test data is collected from six news sites in Taiwan, i.e., *China Times, Commercial Times, China Times Express, United Daily News, Tomorrow Times,* and *China Daily News,* through the Internet. There are in total 17,877 documents (near 13 MB) recorded from January 1, 2001 to January 5, 2001. The total amount of MUs in this test collection is 189,774. After clustering, there are 3,146 events. Because of assessor cost, only 12 events were selected randomly in the first stage. A total of 60 questionnaires (five questions of each event) are made man-

TABLE 2. Results using QA task.

	FULL	BASIC	TFWV	PSWV	TFNV	PSNV
Size (Byte)	59,637	12,974	12,002	12,348	15,192	15,267
Reduction Rate-S	1	0.22	0.20	0.21	0.25	0.26
Reading Time (s)	2,224	780	744	660	816	804
Answering Time (s)	1,752	1,236	1,200	1,128	1,356	1,260
R&A Time (s)	3,976	2,016	1,944	1,788	2,172	2,064
Reduction Rate-T	1	0.51	0.49	0.45	0.55	0.52
Precision	0.923	0.525	0.513	0.519	0.502	0.513
Variance	0.010	0.047	0.095	0.054	0.712	0.061

ually with answers to their related documents. They are all factual questions and some sample questions are shown below. Here $Q_{i,i}$ denotes the *j*th question in event *i*:

(Q10,1) 遠東航空總經理是誰? Who is the general manager of the

Who is the general manager of the Far East Airline Company?

(Q_{2,2}) 新生國小共耗資多少錢 How much did it spend to build the Xin-Sheng primary school?

Moreover, 12 members of our laboratory who are all graduate students majoring in computer science are selected to conduct the following experiments: (1) full text (FULL), (2) the base line system (BASIC) (3) term frequency only with vote strategy (TFWV, i.e., $\lambda = 1$), (4) informative words with vote strategy (PSWV, i.e., $\lambda = 1/2$) (5) term frequency without vote strategy (TFNV, i.e., $\lambda = 1$), and (6) informative words only without vote strategy (PSNV, i.e., $\lambda = 1/2$). Each assessor evaluates a summarization method twice, using different question sets (i.e., answer only once per event).

To evaluate objectively, each assessor does not know the text types what he (she) assesses. The experimental results are shown in Table 2. R&A time means the summation of reading time and answering time. On the one hand, Reduction Rate-S and Reduction Rate-T mean the relative reduction rate of size and R&A time, respectively. The definition of Relative Reduction Rate of size is (Size of a specified system)/(Size of FULL). The average precision and its relative variance of each text type are also given to show the statistical information.

Discussion

Several observations from Table 2 are shown below.

- 1. The size of TFNV and PSNV is larger than that of BASIC (near 15%), but the precision rate of TFNV and PSNV is lower than that of BASIC.
- 2. The size of TFWV and PSWV is smaller than that of BASIC, and their precision rate is also smaller than that of BASIC.
- 3. The precision rates of both TFWV and PSWV are larger than those of TFNV and PSNV.

The above observations are out of our expectation. From observations (1) and (2), the informative words seem not to be useful in MU selection. From observation (3), the vote strategy seems to be useful to improve the precision. In other words, neglecting the news story reported by only one reporter seems to have no problems in Q&A. However, due to limitations and drawbacks of human assessment, evaluation shown below in the QA task may mislead.

- 1. Due to different backgrounds among human assessors, the evaluation is unable to be objective. We have to conduct several evaluations in order to obtain correct and objective results. Nevertheless, this will be cost-effective.
- Fatigue and limited time to work may effect the assessor assessment. That is, the assessors may quit reading or read too fast and miss the useful information to answer the questions. This will effect the precision.
- 3. Due to the high cost of the assessors, the large-scale evaluation is nearly impossible.

An Evaluation Model Using Question Answering System

Model Using Question Answering System

To improve the QA task and verify the experimental results, a Question Answering system is used to substitute the human assessors in Figure 2. The flow of the revised evaluation model is shown in Figure 3. Both full texts and summaries are read by Question Answering systems, and the systems find the answers from full texts and summaries, although the efficiency of a Question Answering system may affect the evaluation results, which is fair for all summarization models under the same evaluation environment.



FIG. 3. Revised evaluation model.

TABLE 3. Results with small-scale data using a question answering system.

	FULL	BASIC	TFWV	PSWV	TFNV	PSNV
Precision of OA Task	0.923	0.525	0.513	0.519	0.502	0.513
Precision of Best-1	0.881	0.441	0.407	0.457	0.475	0.475
Precision of Best-2	0.915	0.475	0.475	0.508	0.576	0.559
Precision of Best-3	0.949	0.491	0.475	0.508	0.576	0.559
Precision of Best-4	0.966	0.508	0.491	0.525	0.576	0.559
Precision of Best-5	0.966	0.541	0.517	0.525	0.576	0.559
OA MRR	0.914	0.493	0.476	0.487	0.508	0.517
Relative MRR	1	0.576	0.521	0.533	0.556	0.566

The Question Answering system we adopted was borrowed from Lin and Chen (1999), whose main strategies are keyword matching and question-focus identifying. This system has been used in open domain question and answering on heterogeneous data (Lin et al., 2001). It is composed of three major modules shown as follows:

- Preprocessing the Question Sentences: At first, the parts
 of speech are assigned to the words in question sentences. Then, stop words are removed. The remaining
 words are transformed into the root forms and considered
 as the keywords of question sentences. For each keyword, we find all synonyms from the related thesaurus,
 for example, Chinese–English WordNet (H.H. Chen,
 Lin, & Lin, 2002). Those terms are the expansion set of
 the keywords. Moreover, no matter whether the keyword
 is a noun, a verb, an adjective or an adverb, all the
 possible morphological forms of the word are also added
 into this set.
- 2. Retrieving the Documents Containing Answers: A full text retrieval system is implemented to decrease the number of documents to be searched for the answering sentences. Each keyword of an expanded question sentence is assigned a weight. Especially, those words tagged with proper-noun markers have been assigned higher weights. This is because they may be the focus of the answer. The score of a document D is calculated according to formula (6).

$$\operatorname{score}(D) = \sum_{t \in D} \operatorname{weight}(t)$$
 (6)

where t is one of the keywords in the expanded question sentence. Those documents, whose scores are larger than a threshold, are selected as the answering documents. Threshold is set to the sum of weights of the words in the original question sentences. If no document scores are larger than the threshold, no answer to the question is reported.

3. *Retrieving the Sentences Containing Answers:* Finally, each sentence in the retrieved documents is examined. Those sentences that contain most words in the expanded question sentence are retrieved. The top five sentences are regarded as the answers. The answers are sorted according to the number of matching words and the retrieving scores computed at step (2).

Evaluation

The experimental results using the same data shown in a previous section are shown in Table 3. The precision from Table 2 is reproduced here for comparison. After the Question Answering system reads all documents of 12 events, it will propose five plausible answers for each question. Mean Reciprocal Rank (MRR) (Voorhees, 2000) used in TREC QA task is adopted:

$$MRR = \sum_{i=1}^{N} r_i / N, \qquad (7)$$

where $r_i = 1/\text{rank}_i$ if $\text{rank}_i > 0$, or 0 if $\text{rank}_i = 0$; rank_i is the rank of the first correct answer of the *i*th question, and N is total number of questions. That is, if the first correct answer is at rank 1, the score is 1/1 = 1; if it is at rank 2, the score is 1/2 = 0.5, and so on. If no answer is found, the score is 0. To compare with the precision of QA task in Table 2, we also use five strategies (e.g., Best-1, Best-2, and so on) to compute the precision of the Question Answering system. Best-*n* means the answer must appear in the first *n* results reported. Furthermore, to show the feasibility of the proposed evaluation method, we also provide a large-scale experiment that will be discussed in the next section, in which case human assessment is impossible.

Discussion

Because the Question Answering system avoids the above limitation and drawback of human assessments, the precisions of some types of summarization text are different from the results shown in Table 2. Observing Tables 2 and 3, there are some differences shown below:

- 1. QA_MRR values of TFNV and PSNV are larger than those of the corresponding TFWV and PSWV. Thus, we can conclude that the vote strategy will lose some useful information.
- QA_MRR values of PSWV and PSNV are larger than those of the corresponding TFWV and TFNV. We can draw the conclusion that using both term frequency and document frequency of informative words will select

	FULL	BASIC	TFWV	PSWV	TFNV	PSNV
Size (Kbyte)	13,137	1,786	1,771	1,773	2,226	2,218
QA_MRR	0.515	0.314	0.342	0.346	0.359	0.380
Relative MRR	1	0.610	0.664	0.672	0.697	0.738

more important MUs than only using term frequency of informative words.

3. Comparing the precisions of QA task with the corresponding precisions of the best-5 strategy, we found that there are some significant differences on the precisions of TFNV and PSNV. The related precisions of the best-5 strategy are obviously higher than those of QA Task (i.e., 0.576 > 0.502 and 0.559 > 0.513, respectively). It shows that the Question Answering system can assess the summarization system more effectively than human assessors, who are easy to neglect or stop reading some contents in summaries. Furthermore, we also found that the values of QA_MRR can also truly reflect the similar tendency to the best n strategy by comparing the related values. Thus, we conclude that the Question Answering system can be used to assess the automatic summarization system in place of human assessors in the large-scale experiments.

Experiments Using Large Documents and Results

Data Set

From the above analysis, we can conclude that a high performance Question Answering system can be used to replace the role of human assessors. Besides the evaluation time and scale, it can obtain more objective and precise results. In the next experiment, the complete data set described earlier is used. We divided all the 3,146 events of that data set to 14 graduate students in our laboratory and asked them to make 10 questionnaires per person by reading the titles of news articles. A total of 140 new questionnaires are made, and among them, 93 questions have answers. Thus, using these practical questions we can further observe the performance of QA system in text summarization evaluation. Some samples of questions are shown below.

(Q68) 英特爾最新發表產品為何 What is the newest product of Intel Company?
(Q95) 歐拉朱萬何時受傷 When was Mr. Olajuwon wounded?

Experimental Results and Discussion

Table 4 shows the experimental results using a large set of documents. Several points are concluded as follows.

- 1. Due to the increase of document size, the QA_MRR of all models decreased.
- 2. Due to increasing noise of FULL, the QA_MRR of

FULL drops drastically. The relative MRRs of the other models increased when comparing with Table 3.

- 3. The QA_MRR values of TFWV, PSWV, TFNV, and PSNV are also larger than the value of BASIC. This is consistent with the above results in small-scale evaluation using the QA system. Thus, informative words in MU's selection presents good performance.
- 4. The QA_MRR values of PSWV and PSNV are also larger than those of TFWV and TFNV, respectively. To achieve better results, it is recommended to use a combination of term frequency and document frequency in MU's selection.
- 5. Because the performance of each model has similar results to those shown in Table 3, it is feasible to use the QA system in evaluating the performance of large-scale multiple document summarization.

Extension to Multilingual News Summarizer

The multidocument summarization architecture can be extended to multilingual multiple news summarization system. However, several more issues shown below have to be addressed:

- 1. *Translation among news stories in different languages:* The basic idea in multiple document summarizations is to identify which parts of news articles present similar reports. Because the news stories are in different languages, some kind of translation is required, for example, term translation. Besides the problem of translation ambiguity, different news sites often use different names to refer the same entity. The translation of named entities, which are usually unknown words, is another problem.
- 2. *Idiosyncrasy among languages:* Different languages have their own specific features. For example, a Chinese sentence is composed of characters without word boundary. Word segmentation is indispensable for Chinese. Besides, Chinese writers often assign punctuation marks at random, how to determine a meaningful unit for similarity checking is a crucial issue. Thus some tasks may be done for specific languages during summarization.
- 3. Implicit information in news reports: Some information is implicit in news stories. For example, the name of a country is usually not mentioned in a news article reporting an event that happened in that country. On the contrary, the country name is important in foreign news. Besides, time zone is used to specify date/time implicitly in the news.



FIG. 4. Architecture of multilingual summarization system.

4. User preference: When users want to read documents in their familiar languages, news fragments in some languages are preferred to those in other language. Machine translation should be the most appropriate method to translate news fragments. Besides, if a user prefers the news from the view of his country or some news sites, we should provide such an option.

Figure 4 shows the architecture of a multilingual summarization system, which is used to summarize the news from multiple sources in different languages. It is composed of three major components: several monolingual news clusterers, a multilingual news clusterer, and a news summarizer.

The monolingual news clusterer receives a news stream from multiple on-line newspapers in its respective language, and directs them into several output news streams by using events. The multilingual news clusterer then matches and merges the news streams of the same event but in different languages in a cluster. The news summarizer summarizes the news stories for each event.

The possible tasks for each component depend on the languages used. Some major tasks of a monolingual clusterer are listed below.

- 1. Identifying word boundaries for Chinese and Japanese sentences
- 2. Extracting named entities like people, place, organization, time, date, and monetary expressions
- 3. Clustering news streams based on predefined topic set and named entities.

The task for the multilingual clusterer is to align the news clusters in the same topic set, but in different languages. It is similar to document alignment in comparable corpus. Named entities are also useful cues.

The major tasks for the news summarizer are shown as follows.

- 1. Partitioning a news story into several meaningful units (MUs)
- 2. Linking the meaningful units, denoting the same thing from different news reports
- 3. Displaying the summarization results under the consideration of language type users prefer, information decay, and views of reporters.

H.H. Chen, Kuo, and Su (2003) proposed translation methods, multilingual clustering models, and visualization in multilingual summarization.

Concluding Remarks

This article presents a multiple Chinese news summarization system. A two-level clusterer is employed to collect news articles for the same event. A summarizer identifies the similar MUs from news articles belonging to the same event, and displays the summarization results based on two strategies. The information decay strategy helps reducing the redundancy, and the user can get all the information provided by the news. But it may still read too much. Also, the order of the sequence is not according to the importance. The user may quit reading and miss the information that follows. The voting strategy gives a shorter summarization, but it also misses some unique information reported by some news sites. Meanwhile, the limitation of the traditional evaluation model is also addressed. Therefore, this article further presents a multidocument summarization system using informative words and introduces a QA system to reduce the evaluation loads. Using the normalized term frequency and document frequency, the informative words can be extracted effectively. The informative words are shown to be more useful to select sentences for generating summaries than the heuristic rule. Moreover, the sentences in the summaries can be ordered according to the total number of informative words. In this way, the important sentences are generated in the early part. Thus, the summaries can be compressed easily by deleting sentences from the end without losing much important information, and the length of summary can be adjusted robustly. On the other hand, the evaluation processes show that QA system can play an important role in conducting large-scale evaluation of multidocument summarization, and makes the results more objective than the conventional ways. Furthermore, to show the feasibility to develop the multilingual clustering summarizer, the proposed architecture for multilingual clustering is sketched and several issues are also discussed.

Acknowledgments

The research in this article was partially supported by National Science Council under the contracts NSC89-2218-E-041 and NSC90-2213-E-002-045.

References

- Barzilay, R., Elhadad, N., & McKeown, K.R. (2001). Sentence ordering in multidocument summarization. In Proceedings of Human Language Technology Conference (HLT-2001), San Diego.
- Brunn, M., Chali, Y., & Pinchak, C.J. (2001). Text summarization using lexical chains. In Proceedings of First Document Understanding Conference, New Orleans, LA.
- Carbonell, J., & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR '98) (pp. 335–336), Melbourne, Australia.
- Chen, H.H. (1994). The contextual analysis of Chinese sentences with punctuation marks. Literal and Linguistic Computing, 9(4), 281–289.
- Chen, H.H., Ding, Y.W., & Tsai, S.C. (1998). Named entity extraction for information retrieval. Computer Processing of Oriental Languages, Special Issue on Information Retrieval on Oriental Languages, 12(1), 75–85.
- Chen, H.H., & Huang, S.J. (1999). A summarization system for Chinese news from multiple sources. In Proceedings of 4th International Workshop on Information Retrieval with Asia Language (pp. 1–7).
- Chen, H.H., Kuo, J.J., & Su, T.C. (2003). Clustering and visualization in a multi-lingual multi-document summarization system. In Proceedings of 25th European Conference on Information Retrieval Research, Lecture Note in Computer Science, 2633 (pp. 391–401), Springer.
- Chen, H.H., Lin, C.C., & Lin, W.C. (2002). Building a Chinese–English WordNet for translingual application. ACM Transactions on Asian Language Information Processing, 1(2), 103–122.
- Chen, K.H., & Chen, H.H. (1995). A corpus-based approach to text partition. In Proceedings of International Conference of Recent Advances on Natural Language Processing (pp. 152–160), Tzigov Chark, Bulgaria.
- Chen, K.H., Huang, S.J., Lin, W.C., & Chen, H.H. (1998). An NTUapproach to automatic sentence extraction for summary generation. In Proceedings of TIPSTER Text Phase III 18-Month Workshop, 4–6 May, Fairfax, VA.
- Edmundson, H.P. (1964). Problems in automatic extracting. Communications of the ACM, 7, 259–263.
- Edmundson, H.P. (1969). New methods in automatic extracting. Journal of the ACM, 16, 264–285.
- Firmin Hand, T., & Sundheim, B. (1998). TIPSTER-SUMMAC Summarization Evaluation. In Proceedings of the TIPSTER Text Phase III Workshop, Washington.
- Fukumoto, F., & Suzuki, Y. (2000). Event tracking based on domain dependency. In Proceedings of 23rd SIGIR (pp. 57–64).
- Goldstein, J., Kantrowitz, M., Mittal, V., & Carbonell, J. (1999). Summarizing text documents: Sentences selection and evaluation metrics. In Proceedings of SIGIR 1999 (pp. 121–128).
- Goldstein, J., Mittal, V., Carbonell, J., & Callan, J. (2000). Multi-document summarization by sentence extraction. In Proceedings of ANLP/ NAACL Workshop on Summarization (pp. 40–48), Seattle, WA.
- Hatzivassiloglou, V., Klavans, J.L., & Eskin, E. (1999). Detecting text similarity over short passages: Exploring linguistic feature combinations via machine learning. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-1999) (pp. 203–212), College Park, MD.
- Hatzivassiloglou, V., Klavans, J.L., Holcombe, R., Barzilay, R., Kan, M.Y., & McKeown, K.R. (2001). SIMFINDER: A flexible clustering tool for summarization. In Proceedings of workshop of automatic summarization (NAACL-2001) (pp. 41–49), Pittsburgh.

- Hovy, E., & Marcu, D. (1998). Automated text summarization. Tutorial in COLING/ACL98.
- Kupiec, J.M., Peterson, J., & Chen, F. (1995). A trainable document summarizer. Proceeding of the 18th Annual International Conference on Research and Development in Information Retrieval (ACM SIGIR '95) (pp. 68–73), Seattle, WA.
- Li, C.N., & Thompson, S.A. (1981). Mandarin Chinese: A functional reference grammar. Berkeley, CA: University of California Press.
- Lin, C.J., & Chen, H.H. (1999). Description of preliminary results to TREC-8 QA task. In Proceedings of the Eighth Text Retrieval Conference (pp. 363–368).
- Lin, C.J., Chen, H.H., Liu, C.J., Tsai, C.H., & Wung, H.C. (2001). Open domain question answering on heterogeneous data. In Proceedings of ACL Workshop on Human Language Technology and Knowledge Management (pp. 79–85), Toulouse, France.
- Lin, C.Y., & Hovy, E. (1997). Identifying topics by position. In Proceedings of the 5th ACL Conference on Applied Natural Language Processing (pp. 283–290), Washington, DC.
- Lin, C.Y., & Hovy, E. (2002). From single to multi-document summarization: A Prototype system and its evaluation. In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL-02), Philadelphia, PA.
- Luhn, H.P. (1958). The automatic creation of literature abstract. IBM Journal of Research and Development, 2(2), 159–165.
- Mani, I., & Bloedorn, E. (1997). Multi-document summarization by graph search and matching. In Proceedings of the 14th National Conference on Artificial Intelligence (pp. 622–628), Providence, RI.
- Mani, I., House, D., Klein, G., Hirschman, L., Firmin, T., & Sundheim, B. (1998). The TIPSTER SUMMAC Text Summarization Evaluation: Final report. Technique Report, Automatic Text Summarization Conference.
- McKeown, K.R., Klavans, J.L., Hatzivassiloglou V., Barzilay, R., & Eskin, E. (1999). Toward multidocument summarization by reformulation: Progress and prospects. In Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-99) (pp. 453–460), Orlando, FL.
- McKeown, K.R., & Radev, D.R. (1995). Generating summaries of multiple news articles. In Proceedings of 18th Annual International Conference on Research and Development in Information Retrieval (ACM SIGIR '95) (pp. 74–82), Seattle, WA.
- Mei, C.J., Zhu, Y.M., Gao, Y.C., & Yin, H.S. (1982). tong2yi4ci2ci2lin2. Shanghai: Shanghai Dictionary Press.
- MUC. (1998). Proceedings of 7th Message Understanding Competition. Available: http://www.muc.saic.com/proceedings/proceedings_index.html
- Radev, D.R., Blair-Goldensohn, S., & Zhang, Z. (2001). Experiments in single and multi-document summarization using MEAD. In Proceedings of first document understanding conference, New Orleans, LA.
- Radev, D.R., & McKeown, K.R. (1998). Generating natural language summaries from multiple on-line sources. Computational Linguistics, 24(3), 469–500.
- Tsutomo, H., Sasaki, T., & Isozaki, H. (2001). An extrinsic evaluation for question-biased text summarization on QA tasks. In Proceeding of Workshop of Automatic Summarization (NAACL-2001) (pp. 61–81), Pittsburgh, PA.
- Voorhees, E. (2000). QA track overview (TREC) 9. [On-line] Available: http://trec.nist.gov/presentations/ TREC9/qa/index.htm
- Yang, Y. (1981). The research on punctuation marks. Hong Kong: Tianjian Publishing Company.

Appendix. Summarization Results

There are two kinds of summarization results, i.e., focusing summarization and browsing summarization. The former lists only those focused. The complete news articles can refer to the browsing summarization part. In the latter, the summarization is highlighted and the redundancy is displayed in gray.

Focusing Summarization

 受芭比絲颱風路徑偏東影響,中央氣象局昨 (廿六)日晚間九時發布全台灣地區陸上颱 風警報,臺灣地區全部為颱風警戒區域,預估 今天上午起南部地區將進入芭比絲颱風暴風 圈內,颱風中心距本島一百五十公里沿臺灣海 峽北上,明(廿八)天上午以後暴風圈才會逐 步脫離本島。

(Due to the effect of typhoon Babs route across the east, the Central Weather Bureau declared a land typhoon alarm at 9:00 p.m. last (26th) night. The whole Taiwan area is under the typhoon detective area, the southern area is forecasted to be in typhoon Babs storm circle. The typhoon center is 150 km from the island and is along the top north of the Taiwan straits, the storm circle will leave the island only after tomorrow (28) morning.)

 中央氣象局昨日晚間針對臺灣本島、澎湖
 金門、馬祖等地區及各海域發出輕度颱風芭 比絲的海上、陸上颱風警報,由於芭比絲颱風
 外圍環流挾帶著豐沛的水氣直撲臺灣西南部 在沒有中央山脈屏障的情況下原先僅出現
 於臺灣東半部、北部地區的豪大兩將擴及全 省氣象局特別呼籲西南部沿海低漥地區應嚴
 防海水倒灌。

(The Central Weather Bureau declared a light typhoon alarm on sea and land for Taiwan, Penghu, Kinmen, Mazu, and each sea area. As typhoon Babs outer circumfluence carries plentiful mist, striking towards Taiwan's southwest side. As there is no shield from central mountain chain, the heavy rains that were supposed to show up only in the east half and northern area will expand to the whole providence. The central weather bureau especially called upon the lower-laying lands at the southwest along the coast to take strict precautions against the sea water to flow backwards.)

- 3. 中央氣象局指出芭比絲颱風暴風圈接 觸廣東陸地後強度已在昨天下午二時減弱 為輕度颱風芭比絲颱風昨晚位於澎湖西 南方二百六十公里海面上預測以時速十三 公里朝向北北東轉東北移動,中心附近最大 風速每秒三十公尺相當於十一級風 瞬間最大陣風每秒四十公尺相當於十三級風 七級風暴風半徑二五〇公里預測今天晚間 八時的中心位置在金門東方一〇〇公里之海 面上。
 - (The Central Weather Bureau indicated that after typhoon Babs landed on the Canton land, its strength has weakened to a light typhoon at 2:00 p.m. yesterday. Typhoon Babs was on the sea at 260 km southwest of Penghu last night, and is predicted to be moving towards the northeast side at a speed of 13 km. The highest wind speed near the center is 30 seconds per meter, which is equal to a 11th degree wind (a violent gale), the highest wind speed 40 seconds per meter, which is equal to a 13 degree wind (a moderate gale), force 7 wind with a storm radius of 250 km, its center is predicted to be at 100 km east of Kinmen at 8:00 pm tonight.)
- 4. 首當其衝的澎湖和金門今日清晨即 已進入暴風圈內並出現十二級的最大陣風。 (Penghu and Kinmen the first to be affected have already entered in the storm circle this morning and have a maximum wind of 12 degrees. (hurricane))
- 5. 根據天氣資料顯示臺中以南及花蓮以南 地區的風雨自昨日晚間起已經愈來愈強特別 是因為西南部地區沒有山脈屏障氣象局預測今 日午後平地地區一般風力可達五到六級 陣風可達九到十級濱海地區風力則有六到七級 陣風可達十一級以上各地的風力及雨勢都將 相當驚人。

(According to the weather data, the wind and rain at the south of Taichung and Hwalian have been stronger since yesterday. Especially since there is no shield from the mountain in the southwest area. The Central Weather Bureau predicts the general wind power to be level 5 to 6 (strong breeze), gust wind up to level force nine to 10 wind (a strong gale to a whole gale), wind power by the sea is about 6 to 7 (a strong breeze to a moderate gale), gust wind higher the force 11 (a storm), the wind power and rain are all very alarming everywhere.)

Browsing Summarization

