

Toward the “At-a-glance” Summary: Phrase-representation Summarization Method

Yoshihiro UEDA, Mamiko OKA, Takahiro KOYAMA and Tadanobu MIYAUCHI
Industry Solutions Company, Fuji Xerox, Co., Ltd.
430 Sakai, Nakai-machi, Kanagawa 259-0157, JAPAN

{Ueda.Yoshihiro, oka.mamiko, koyama.Takahiro, miyauchi.Tadanobu}@fujixerox.co.jp

Abstract

We have developed a summarization method that creates a summary suitable for the process of sifting information retrieval results. Unlike conventional methods that extract important sentences, this method constructs short phrases to reduce the burden of reading long sentences. We have developed a prototype summarization system for Japanese. Through a rather large-scale task-based experiment, the summary this system creates proved to be effective to sift IR results. This summarization method is also applicable to other languages such as English.

Introduction

Summaries are used to select relevant information from information retrieval results. The goal of summarization for such “indicative” use is to provide fast and accurate judgement.

Most automatic summarization systems adopt the “sentence selection” method, which gives a score to every sentence on the basis of its characteristics, such as word frequency, the position in which it appears, etc. and selects sentences with high scores.

The sentences collected in such a way tend to be so long and complex that the reader must reconstruct the structure while reading them. Reading such sentences involves some annoyance.

Our aim is to reduce this burden by providing an “at-a-glance” summary.

Phrase-representation summarization is a method to create the “at-a-glance” summary for the Japanese language. Here we present the concept, the algorithm, and evaluation of the efficacy of the summary produced by a prototype based on this method. Extension to English is also discussed.

1 The Concept

Examples of an “at-a-glance” summary are the headlines of news articles. The headline provides information for judging whether the

article is to be read or not and, in this sense, it is really “indicative.” The characteristics are:

- Brevity (short in length)
- Simplicity (less embedded sentences)

We use “phrases” to represent the simplicity characteristic¹ and set our goal to create phrase-represented summaries, which provide the reader with an outline of the document, avoiding reading stress by enumerating short phrases containing the important words and concepts composed from these words.

The method we adopted to achieve this goal is to **construct** such **phrases** from the relations between words rather than extracting important sentences from the original document.

2 Summarization Method

2.1 Outline of the Algorithm

Here we give a short description of the outline of this method using the example shown in Fig. 1.²

¹ The word “phrase” used here is not of the linguistic sense but an expression for “short” and “simple.” In Japanese, there is no rigid distinction between “phrase” and “clause.”

² In this paper, Japanese words are represented in English as much as possible. The words left in Japanese are shown in italics, such as “*ga*” (a particle for AGENT), “*jidai*” (“era”), etc. Each relation name is constructed from a Japanese particle and its function (shown as a case name or an equivalent English

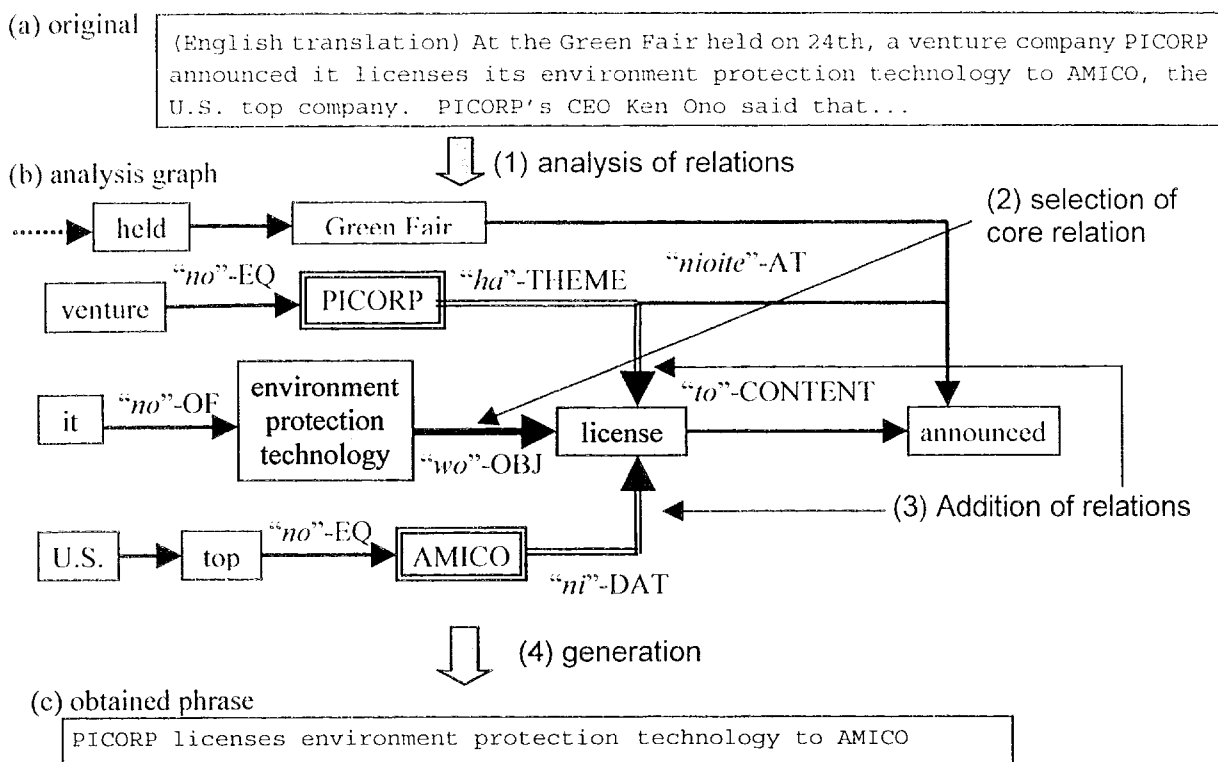


Fig.1: Outline of phrase-representation summarization

The method consists of the following four major steps:

- (1) Syntactic analysis to extract the relations between words
- (2) Selection of the core relation
- (3) Adding relations necessary for the unity of the phrase's meaning
- (4) Generating the surface phrase from the constructed graph

First, the sentences in the given document are analyzed to produce directed acyclic graphs (DAGs) constructed from relation units, each of which consists of two nodes (words) and an arc (relation between the words). Each node is not only a single word but also can be a word sequence (noun group).

Then an important relation is selected as a "core" relation. In Fig. 1, the arc connecting the two shaded nodes is selected as the "core."

The core relation alone carries insufficient information to convey the content of the original document. Additional arcs (represented by

preposition).

double lines) are attached to narrow the information the phrase supplies.

The following short phrase can be generated from the selected nodes and arcs in the graph:

PICORP licenses (its) environment protection technology to AMICO.³

Phrase-representation summarization enumerates such short phrases to give the readers enough information to grasp the outline of a document. This algorithm is explained in the next section.

2.2 Further description of each step

The steps shown in the previous section consists of a cycle that produces a single phrase. The cycles are repeated until the generated phrases satisfy a predefined condition (e.g. the length of the summary). The scores of the words used in the cycle are reduced by a predefined cut-down

³ This short sentence can be expressed as a phrase – in the linguistic sense – in English:

PICORP's licensing (its) environment protection technology to AMICO.

ratio to avoid frequent use of the same words in the summary.

The basic algorithm is shown in Fig. 2.

Relation Analysis

Syntactic analysis is applied to each sentence in the document to produce a DAG of the relations of words. We use a simple parser based on pattern matching (Miyauchi, et al. 1995), one of whose rules always judges each case dependent on its nearest verb. Some of the misanalysis will be hidden by “ambiguity packing” in the “additional relation attachment” step.

Relation Scoring

An importance score is provided for each relation unit (two nodes and an arc connecting them).

First, every word is scored by its importance. This score is calculated based on the $tf \cdot IDF$ value (Salton, 1989)⁴.

Then, the relation score is calculated as follows:

$$\text{Score} = S_{rel} * (W1 * S1 + W2 * S2)$$

Here, $S1$ and $S2$ are the scores of the two words connected by relations. The score of a word sequence is calculated by decreasing the sum of the scores of its constituent words according to the length of the word sequence.

$W1$ and $W2$ are the weights given to each word. Currently, all words are equally treated ($W1 = W2 = 1$).

S_{rel} is the importance factor of the relation. The relations that play central roles in the meaning, such as verb cases, are given high scores, and the surrounding relations, such as “AND” relations, are scored low. The relation scores for modifier-modified relations such as adverbs are set to 0 to avoid selecting them as the core relations.

Core relation selection

The relation unit with the highest score among all relations is selected as the “core relation.”

Additional relation attachment

The information that the core relation carries is usually insufficient. Additional relations are attached to make the information the phrase

⁴ IDF is calculated from 1 million WWW documents gathered by a Web search engine.

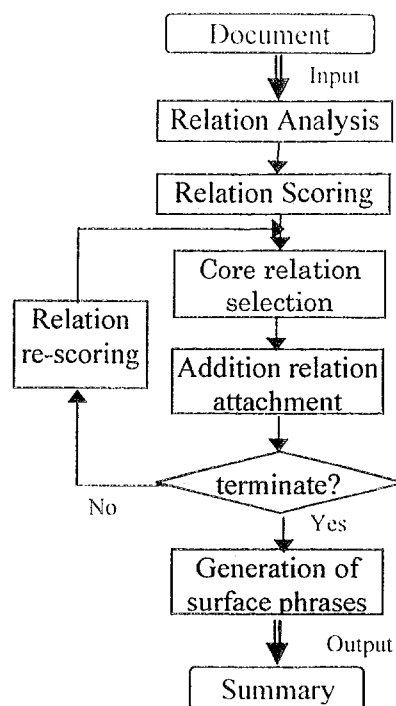


Fig.2 Basic flow of the algorithm

supplies more specific and to give the reader sufficient information to infer the content of the original document. The following relations are a part of the relations to be attached.

● Mandatory cases

Relations that correspond to mandatory cases are attached to verbs. Mandatory case lists are defined for verbs except for those that share the common mandatory case list, which includes “*ga*”-AGENT, “*wo*”-OBJ and “*ni*”-DATIVE. “*Ha*”-THEME, “*mo*”-ALSO, and null-marker relations are also treated as mandatory, because they can appear in place of the mandatory relations.

Ex.) AMICO “*ga*”-AGENT release
 → AMICO “*ga*”-AGENT
 PDA “*wo*”-OBJ release
 (AMICO releases PDA.)

● Noun modified by a verb

In Japanese, the “verb - noun” structure represents an embedded sentence, and the noun usually fills some gap in the embedded sentence. If the verb in the core relation (noun – verb) consists of such a verb – noun relation, the modified noun is also assumed to carry important information, even if it does not fill the mandatory case (though the case is not

analyzed in the current algorithm). Thus the verb - noun relation is attached to the core.

Ex.) PDA "wo"-OBJ release
→ PDA "wo"-OBJ release
 0-THAT⁵ AMICO
 (AMICO that releases PDA)
PDA "wo"-OBJ release
→ PDA "wo"-OBJ release
 0-THAT plan
 (a plan to release PDA)

● Ambiguity packing

The analysis trees often contain errors because the pattern-base parser doesn't resolve ambiguities. For example, the structure

V 0-THAT N1 "no"-OF N2 (Ving N1's N2)
is ambiguous in Japanese (V can modify either N1 or N2 but the parser always analyzes N2 as modified). If the V-N1 relation is selected as the core, the N1-N2 relation is always attached to the core to include the possible V-N2 relation.

● Modifiers of generic nouns

The concepts brought by generic nouns such as "mono" (thing), "koto" ("that" of that-clause), "baai" (case), "jidai" (era) are not so specific that they usually accompany modifiers to be informative. Here such modifiers are attached to make them informative.

Ex.) era "ni"-TIME emerge
→ confusion "no"-OF era
 "ni"-TIME emerge
 (emerged in the era of confusion)

Termination condition

Judges whether the summaries created so far are sufficient. Currently the termination condition is defined by either the number of produced phrases or the total summary length.

Re-scoring of relations

If the condition is not fulfilled, these steps from selection of the core relation must be repeated to create another phrase. Before selecting a new core, the scores of the words used in this cycle are reduced to increase the possibility for other words to be used in the next phrase. Score reduction is achieved by multiplying the predefined cut-down ratio R ($0 < R < 1$) by the scores of the words

used. Relation scores are re-calculated using the new word scores.

Generation of surface phrases

This process produces DAGs each of which consists of one core relation and several attached relations. In Japanese, the surface phrases can be easily obtained by connecting the surface string of the nodes in their original order. See Chapter 5 for the generation method for English.

3 The Prototype

We developed a prototype of the summarization system based on this algorithm. The development language is Java and the system is working on Windows 95/98/NT and Solaris 2.6⁶.

The time consumed by summarization process is in proportion to the text length and it takes about 700 msec to generate a summary for an A4 sized document (2000 Japanese characters) using a PC with a Celeron processor (500 MHz). Over 95% of the time is consumed in the relation analysis step.

4 Evaluation

We have conducted an experiment to evaluate the system. This section is a short summary of the experiment reported in (Oka and Ueda, 2000).

The aim of a phrase-represented summary is to give fast and accurate sifting of IR results. To evaluate whether the aim was achieved, we adopted a task-based evaluation (Jing, et al. 1998, Mani, et al. 1998). One of the problems of those experiments using human subjects as assessors is inaccuracy caused by the diversity of assessment. To reduce the diversity, first we assign 10 subjects (experiment participants) for each summary sample. The number of subjects was just 1 or 2 in the previous task-based experiments. Second, we gave the subjects a detailed instruction including the situation that led them to search the WWW.

4.1 Experiment Method

The outline of the evaluation is as follows:

⁵ "0" shows that there are no particles or any other words connecting two words. Japanese doesn't require anything like relative pronouns.

⁶ Java and Solaris are the trademarks of Sun Microsystems. Windows and Celeron are the trademarks of Microsoft and Intel, respectively.

- Assume an information need and make a query for the information need
- Prepare simulated WWW search results with different types of summaries; (A) first 80 characters, (B) important sentence selection (Zechner, 1996), (C) phrase-represented summary, (D) keyword enumeration. The documents in the simulated search result set are selected so that the set includes an appropriate number of relevant documents and irrelevant documents.
- Have subjects judge from the summaries the relevance between the search results and the given information need. The judgement is expressed in four levels (from higher to lower: L3, L2, L1, and L0, which is judged to be irrelevant).
- Compare the relevance with the one that we assumed.

The documents the user judges to be relevant compose a subset of the IR results and it should be more relevant to the information need than the IR results themselves. Because we have introduced three relevance levels, we can assume three kinds of the subsets; L3 only, L3+L2, and L3+L2+L1. The subset composed only from the documents with L3 judgement should have a high precision score and the subset including L1 documents should get a high recall score.

4.2 Result

Because recall and precision are in a trade-off relation, here we show the result using f-measure, the balanced score of the two indexes.

$$F - measure = \frac{2 * precision * recall}{precision + recall}$$

The f-measure averages of the experiment result of three different tasks are shown in Fig. 3. It shows that the phrase-represented summaries (C) are more suitable for sifting search results than any other summaries in all cases.

4.3 Discussion

The result can be explained using the number of summaries that contain clues to the information need. Summaries consisting of short units (phrases (C) and keywords (D)) are gathered from the wide range of the original text and accordingly have many chances to include the clues. The actual average numbers of summaries that

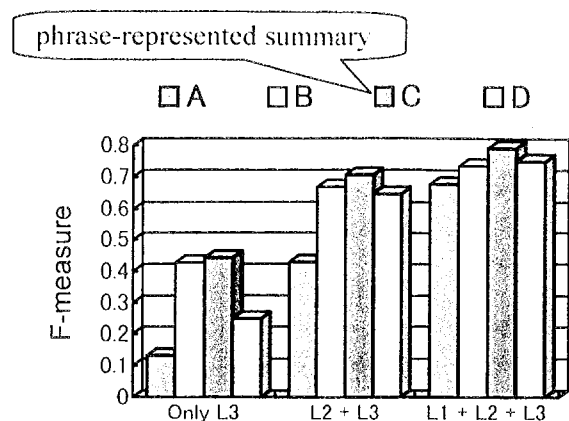


Fig.3 Experiment result

contain the clues are 2.0, 4.3 and 4.7 for (B) sentence, (C) phrases and (D) keywords, respectively. In spite that (D) keywords include more clues than any other samples, they don't get a good f-score. The reason is considered to be due to the lack of information about the relations among keywords.

5 Applicability to Other Languages

Although this algorithm was first developed for the Japanese language, the concept of phrase-representation summarization is also applicable to other languages. Here we show the direction toward its extension to English.

English has a clear concept of "phrase," and simply connected words do not produce well-formed phrases. This requires semantic analysis and generation from the semantic structure.

We will consider the following example again.

Ex.) A venture company PICORP announced to license their environment protection technology to AMICO, a U.S. top company.

If "PICORP" and "license" must be included in the summary and "announce" is not so important, "PICORP license(s)" is the core of the desired phrase. Generating it requires subject resolution of "license" and thus semantic level analysis is required. Moreover, predicate-argument structures are preferable to syntactic trees because the subject and the object are represented in the same level. Unification grammar frameworks such as LFG (Kaplan and Bresnan, 1982) and HPSG (Pollard and Sag, 1994) fulfill these requirements. Fig. 4 is a part of the analysis result represented in LFG.

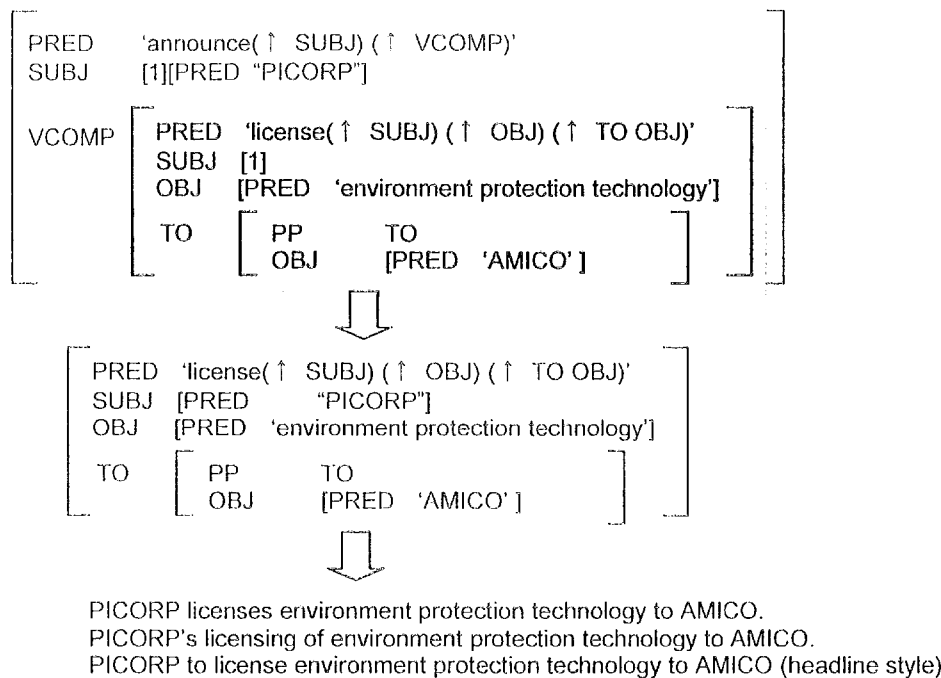


Fig. 4: Analysis and generation of summary

A score is calculated for each feature structure and the core feature structure will be selected by its score instead of selecting a core relation and attaching mandatory relations. In the core feature structure, index [1] is replaced by SUBJ of the top feature structure.

Generating phrases from the feature structure requires templates⁷. Several patterns can be selected to generate phrases:

V-ing (gerund) form

ARG1's PRED-ing ARG2 to ARG3

noun form

ARG1's noun (PRED) of ARG2 to ARG3

to-infinitive form

For ARG1 to PRED ARG2 to ARG3

In this case, the noun form "PICORP's license of the protection technology to AMICO" is avoided because the noun "license" lacks the meaning of "action" or "event." Other rules specific to headlines such as "to-infinitive represents future" can also be introduced.

6 Related Work

Most summarization studies (including Zechner, 1996) are based on important sentence selection and seek better selection methods. We have

pointed out that summaries made by this method tend to be burdensome to read, and have proposed phrase-representation summarization as an alternative. The following studies bear some relation to our study.

The summarization method by Boguraev and Kennedy (1997) adopts "phrasal expression" rather than sentences or paragraphs. However, it begins to create a phrase not from a core relation but a core word (in their words, "topic stamp") and produces multiple phrases containing the same core word; it is therefore not suitable for summaries for sifting IR results. In addition, because it does not consider the roles and importance of the attaching arcs when enriching the core, less important words are often attached to the core. They aimed at supporting fast reading rather than sifting IR results.

Some studies are similar to ours in that they make sentences short. Wakao, et al. (1998) and Mikami, et al. (1998) aim to create closed captioning from an announcer's manuscript by paraphrasing and removing modifiers. This method doesn't remove the "trunk" of the analysis tree and the summaries cannot be made as short as in phrase-representation.

Nagao, et al. (1998) also proposed a method to create summarization based on the relations

⁷ Generation of articles is left to be considered.

between words. They utilize GDA (Global Document Annotation), a tag set that the document author inserts into the document and that contains linguistic information such as sentence structures and reference information. Although this method is similar to ours in some points, the summary consists of sentences and thus does not have “at-a-glance” capability. Most of all, the expectation that every document is tagged linguistically will not be fulfilled until special editors with automatic linguistic tagging become popular.

Conclusion

We introduced the concept of “at-a-glance” summary and showed an algorithm of phrase-representation summarization as a realization of the concept. An experiment shows that the summaries are effective for sifting IR results.

We continue to fine-tune the prototype for further efficacy.

Acknowledgement

We would like to thank our laboratory members who give us valuable suggestions and participated in the experiment.

References

- Bougraev, B. and Kennedy, C. (1997): “Salience-based Content Characterisation of Text Documents,” *Proc. Intelligent Scalable Text Summarization*, pp. 2-9.
- Jing, H., Barzilay, R., McKeown, K. and Elhadad, M. (1998): “Summarization Evaluation Methods: Experiments and Analysis.” In *Intelligent Text Summarization*. pp. 51-59. AAAI Press.
- Kaplan, R. M. and Bresnan, J. (1982): “Lexical-Functional Grammar: A Formal System for Grammatical Representation,” in Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*, MIT Press.
- Mani, I., House, D., Klein, G., Hirschman, L., Obrst, L., Firmin, T., Chizanowski, M., and Sundheim, B. (1998): “*The TIPSTER SUMMAC Text Summarization Evaluation*.” Technical Report MTR 98W0000138, MITRE Technical Report.
- Mikami, M., Yamazaki, K., Masuyama, S. and Nakagawa, S. (1998): “Summarization of News Sentences for Closed Caption Generation,” *Proc. Workshop Program The 4th Annual Meeting of The Association for Natural Language Processing*, pp. 14-21 (in Japanese).
- Miyauchi, T., Oka, M. and Ueda, Y. (1995): “Key-relation technology for text retrieval.” *Proc. the SDAIR '95*, pp. 469-483.
- Nagao, K. and Hasida, K. (1998): “Automatic Text Summarization Based on the Global Document Annotation,” *Proc. COLING-98*, pp. 917-921.
- Oka, M. and Ueda, Y. (2000): “Evaluation of Phrase-representation Summarization based on Information Retrieval Task,” *Proc. ANLP/NAACL 2000 Workshop on Automatic Summarization*, pp. 59--68.
- Pollard, C. and Sag, I. A. (1994): *Head-Driven Phrase Structure Grammar*, The University of Chicago Press.
- Salton, G. (1989): *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley.
- Wakao, T., Ehara, T. and Shirai, K. (1998): “Automatic Summarization for Closed Caption for TV News,” *Proc. Workshop Program The 4th Annual Meeting of The Association for Natural Language Processing*, 7-13 (in Japanese).
- Zechner, K. (1996): “Fast Generation of Abstracts from General Domain Text Corpora by Extracting Relevant Sentences.” *Proc. COLING-96*, pp. 986-989.