A Relevance Feedback Model for Fractal Summarization

Fu Lee Wang[1] and Christopher C. Yang[2]

[1] Department of Computer Science, City University of Hong Kong,
Kowloon Tong, Hong Kong SAR, China
flwang@cityu.edu.hk
[2] Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China
yang@se.cuhk.edu.hk

Abstract. As a result of the recent information explosion, there is an increasing demand for automatic summarization, and human abstractors often synthesize summaries that are based on sentences that have been extracted by machine. However, the quality of machine-generated summaries is not high. As a special application of information retrieval systems, the precision of automatic summarization can be improved by user relevance feedback, in which the human abstractor can direct the sentence extraction process and useful information can be retrieved efficiently. Automatic summarization with relevance feedback is a helpful tool to assist professional abstractors in generating summaries, and in this work we propose a relevance feedback model for fractal summarization. The results of the experiment show that relevance feedback effectively improves the performance of automatic fractal summarization.

1 Introduction

There is an increasing need for automatic summarization in the wake of the recent information explosion. Many summarization models have been proposed, of which the fractal summarization model is the first to apply the fractal theory to document summarization [29][30]. This model generates a summary by a recursive deterministic algorithm that is based on the iterated representation of a document. Traditionally, automatic summarization systems have extracted sentences from the source document according to predefined rules [7][18], but rule-based extraction does not function properly in some extreme cases, and cannot truly reflect the actual circumstances of every individual document. For example, the thematic feature of summarization systems does not always reflect the significance of a term accurately. In addition, the position of key sentences varies from document to document, and can be very difficult for a summarization system to detect. As a result of these problems, the quality of the summaries that are generated by the extraction of sentences by machine is not high.

In most cases, ordinary users will be satisfied with reading the sentences of a document that have been extracted by machine, as it can help them to decide whether a document is useful. Ordinary users are also reluctant to provide explicit feedback, and therefore most summarization techniques are fully automated. Automatic

Z. Chen et al. (Eds.): ICADL 2004, LNCS 3334, pp. 368–377, 2004.

summarization by sentence extraction is desirable due to the large volume of information that is available today, but the quality of the summaries that are generated by fully automated summarization systems is not good enough. Human professionals are the best summarizers because of their incredible summarization capabilities, and thus automatic summarization systems are best employed as a tool to aid summarization. This entails the extraction of important sentences from the source document by machine, and the synthesis of the summary by a human being based on the sentences that have been extracted [5][6]. With such a method of summarization, relevance feedback models are very helpful, because they provide a mechanism for the abstractor to specify the content that is important and that needs to be extracted, and also what is irrelevant and should be excluded. A relevance feedback model for fractal summarization is proposed in this work, and relevance feedback for each summarization feature is also discussed.

The rest of this paper is organized as following. Section 2 reviews the techniques of automatic text summarization. Section 3 presents a fractal summarization model, Section 4 proposes a relevance feedback model for fractal summarization, and Section 5 discusses the results of the experiment. Section 6 provides some concluding remarks.

2 Traditional Automatic Summarization

Traditional summarization models consider a document as a sequence of sentences, and the traditional method of automatic text summarization involves the selection of sentences from the source document based on their significance to the document as a whole [7][18] without consideration of the hierarchical structure of the document. The selection of sentences is based on the salient features of the document, with thematic, location, heading, and cue phrase features being the most widely used summarization features.

- The *thematic feature* was first identified by Luhn [18]. Edmundson proposes the assignment of a thematic weight to keywords that is based on term frequency, and allocates a sentence thematic score that is the sum of the thematic weights of the constituent keywords [7]. The *tfidf* (Term Frequency, Inverse Document Frequency) score is the score that is most widely used to calculate the thematic weight of keywords [22].
- The significance of a sentence can be indicated by its *location* [2] based on the hypotheses that topic sentences tend to occur at the beginning or the end of a document or paragraph [7]. Edmundson proposes the assignment of positive scores to sentences according to their ordinal position in the document, which is known as a location score.
- The *heading feature* is based on the hypothesis that the author conceives the heading as circumscribing the subject matter of the document [7]. A heading glossary is a list that consists of all of the words that appear in the headings and subheadings that have positive weights. The heading score of a sentence is calculated by the sum of the heading weights of its constituent words.
- The *cue phrase feature* that is proposed by Edmundson [7] is based on the hypothesis that the probable relevance of a sentence is affected by the presence of

pragmatic words. A pre-stored cue dictionary of terms with cue weights is used to identify the cue phrases, and the cue score of a sentence is calculated by the sum of the cue weights of its constituent terms.

Typical summarization systems select a combination of summarization features [7][15][17], and the total sentence score ($W_{sentence}$) is calculated as the weighted sum of the scores that are computed by each of the features, for example,

$$W_{sentence} = a_1 \times w_{thematic} + a_2 \times w_{location} + a_3 \times w_{heading} + a_4 \times w_{cue},$$

where $w_{thematic}$ is the thematic score, $w_{location}$ is the location score, $w_{heading}$, is the heading score, and w_{cue} the cue score of the sentence, and a_1, a_2, a_3, and a_4 are positive integers that adjust the weighting of the four features. Sentences with a sentence score that is higher than a given threshold value are selected for the summary. It has been proved that the weighting of the different summarization features has no substantial effect on the average precision [15]. Thus, in our experiment, the maximum score for each feature is normalized to 1, and the total score of each sentence is calculated as the sum of the scores of all of the summarization features without weighting.

3 Fractal Summarization

Of the many automatic summarization models that were proposed in the past, none of was developed based entirely on document structure, and none took into account the fact that human abstractors extract sentences according to the hierarchical structure of a document. The structure of a document can be described as a fractal [26][29][30]. In the past, fractal theory was widely applied in the area of digital image compression, which is similar to text summarization in the sense that both techniques involve the extraction of the most important information from a source and the reduction of the complexity of the source. The fractal summarization model represents the first effort to apply the fractal theory to document summarization [29][30], in which a summary is generated by a recursive deterministic algorithm that is based on the iterated representation of a document.

Many studies of the human abstraction process have shown that human abstractors extract topic sentences according to the structure of a document from the top level to the bottom level until they have extracted sufficient information [8][9]. Advanced summarization techniques take document structure into consideration to compute the probability that a sentence should be included in the summary, but most traditional automatic summarization models consider the source document as a sequence of sentences and ignore its structure. By contrast, fractal summarization generates a summary that is based on the hierarchical document structure [29][30].

The fractal summarization model is based on fractal theory [19], and applies the techniques of fractal view [14] and fractal image compression [1][12]. In fractal image compression, an image is evenly segmented into sets of non-overlapping square blocks, which are known as range blocks, and each range block is subdivided into sub-range blocks until a contractive map can be found that represents the sub-range block. In fractal text summarization, the original document is partitioned into range blocks according to the document structure, which is represented as a fractal tree structure (Figure 1). The important information is captured from the source document by exploring the hierarchical structure and the salient features of the document, and the summarization system then computes the number of sentences to be extracted based on the compression ratio. The system then assigns the number of sentences to the root as the quota of sentences, and the fractal value of the root node is taken to be 1. The system then calculates the sentence score for each sentence in each range block using traditional summarization methods, and the fractal values are propagated to the range blocks according to the sum of the sentence scores. The sentence quota is then shared out among the range blocks according to their fractal value. The system repeats the procedure for each range block to allocate the sentence quota to the sub-range blocks recursively until the quota that is allocated is less than a given threshold value, and the sub-range block is then transformed into key sentences by traditional summarization methods.

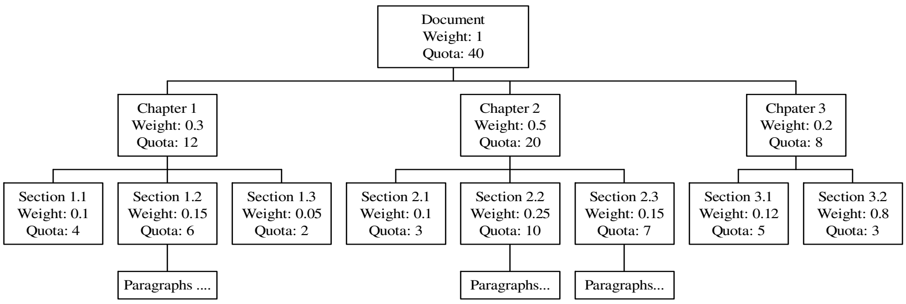

Fig. 1. An Example of a Fractal Summarization Model

The fractal value of each range block is calculated using traditional extraction features. However, traditional extraction features consider the document as a sequence of sentences, and thus are not entirely compatible with the fractal structure. We present a method for the modification of the extraction features to allow them to fully use the fractal structure of a document.

- The *tfidf* score is the most widely used thematic feature approach, but it does not consider document structure. Most researchers assume that the weight of a term remains the same over the entire document, but Hearst claims that a term carries different weights in different locations in a full-length document [11]. We define the i^{th} term in a document as t_i. In fractal summarization, the *tfidf* of the term t_i in a range block is defined as the term frequency within that range block inverse the frequency of the range block that contains the term, i.e.,

$$w_{ir} = tf_{ir} \times \log_2\left(\frac{N' \times |t_i|}{n'}\right),$$

where tf_{ir} is the frequency of term t_i in range block r, N' is the number of range blocks in the document, n' is the number of range blocks that contain the term t_i in the document, and $|t_i|$ is the length of the term t_i. The Fractal Sentence Thematic Score, FSS_T, of the k^{th} sentence (s_k) in range block r is calculated as the sum of the modified *tfidf* score, w_{ir}, of the constituent terms t_i of the sentence s_k, i.e.,

$$FSS_T(k,r) = \sum_{t_i \in s_k} w_{ir} \ .$$

The Fractal Thematic Score FTS_T of range block r is calculated as the sum of the Fractal Sentence Thematic Score FSS_T of all of the sentences in range block r, i.e.,

$$FTS(r) = \sum_{s_k \in r} FSS_T(s_k, r) \ .$$

- Traditional summarization methods assume that the location score of a sentence is static, but fractal summarization calculates the location score based on the document level that is being looked at. A location score is assigned to range blocks according to their position by traditional methods, and the sentences that are inside the range block are hidden. In the fractal summarization model, we consider the location score at the level below the level that is being looked at. The Fractal Location Score FLS of range block r is calculated as the reciprocal of the minimal distance of the range block r to the first sibling range block or last sibling range block under the same parent, i.e.,

$$FLS(r) = \frac{1}{\min(d(r,\ \textit{first sibling of } r), d(r,\ \textit{last sibling of } r))} \ ,$$

  where $d(r, x)$ is the distance function that calculates the number of range blocks between range block r and range block x, inclusively.
- At different abstraction levels, some headings are hidden and some are emphasized. For example, at document level, only the document heading is considered, but if we look at the chapter level, then both the document heading and the chapter heading are considered, and the latter is more important because the main concept of the chapter is represented by its heading. Therefore, the significance of the heading is inversely proportional to its distance. The propagation of fractal value [14] is a promising approach for the calculation of the heading scores for a sentence. If a summarization is being conducted at the internal node x (range block x) with m_x child nodes, then there is a unique path that connects node x to the document root. If the sentence z under the branch of node x contains a term t_i that appears in the heading of node y in the path from the root to node z, then the Fractal Sentence Heading Score FSS_H of sentence k should be the weights w_{ij} of the term t_i divided by the product of the degree of nodes m_{node} in the path from node y to node x, i.e.,

$$FSS_H(k,r) = \sum_{y \in \textit{path from root to } x} \frac{\sum_{t_i \in y \cap s_k} w_{iy}}{\prod_{i \in \textit{path from } y \textit{ to } x} m_i} \ .$$

  The Fractal Heading Score FHS of range block r is calculated as the sum of the Fractal Sentence Heading Score FSS_H of all of the sentences in range block r, i.e.,

$$FHS(r) = \sum_{s_k \in r} FSS_H(s_k, r) \ .$$

- When human abstractors extract sentences from a text, they pay more attention to range blocks with headings that contain bonus word such as "conclusion", because they consider them to be more important parts of the document, and thus contain more sentences that should be extracted. In a document tree, the heading of each range block is examined, and its quota is adjusted accordingly. The Fractal Cue Score *FCS* of range block *r* is calculated as the sum of the cue weight of all the terms in the heading of that range block, i.e.,

$$FCS(r) = \sum_{t_i \in heading(r)} cue(t_i),$$

where *heading*(*r*) is the heading of range block *r* and *cue*(t_i) is the cue weight of term t_i.

In our system, the maximum score for each feature is normalized to 1, and the Range Block Significance Score *RBSS* of a range block is calculated as the sum of the normalized scores of each feature for that range block, i.e.,

$$RBSS(r) = NFTS(r) + NFHS(r) + NFLS(r) + NFCS(r),$$

where *NFTS*, *NFHS*, *NFLS*, *NFCS* are the normalized fractal thematic, heading, location, and cue score, respectively, of range block *r*. The fractal value F_v of the root of the document is 1, which is propagated to the child nodes according to the formula

$$Fv(child\ \ of\ \ x) = C \bullet Fv(x) \left(\frac{RBSS\ (child\ \ of\ \ x)}{\sum_{y \in children\ \ of\ \ x} RBSS\ (y)} \right)^{-1/D}.$$

C is a constant between 0 and 1 that controls the rate of decay and *D* is the fractal dimension, both of which are taken as 1, identically to fractal view experiment [14]. The sentence quota of a summary is calculated based on the compression ratio, and is shared among the child nodes according to their fractal values. The optimal length of a summary that is generated by the extraction of a fixed number of sentences is 3 to 5 sentences [10], and thus if the quota of a node exceeds the default threshold value of 5 sentences, then it will be propagated to grandchild nodes iteratively.

We conducted an experiment that compared the fractal summarization and traditional summarization of annual reports in Hong Kong [29][30], and found that fractal summarization produces a summary with a wider coverage of information subtopics than traditional summarization. A user evaluation by ten participants was conducted to compare the performance of the fractal summarization and the traditional summarization which doesn't consider the hierarchical document structure. The results show that all of the participants considered the summary that was generated by the fractal summarization method to be the better summary. The fractal summarization method achieved a precision of up to 91.3% and of 87.1% on average, but the traditional summarization only achieved a maximum precision of 77.5% and an average of 67.0%. The results also show that fractal summarization outperforms traditional summarization at a 99.0% confidence level.

4 Relevance Feedback

Many studies have shown that relevance feedback can greatly improve the performance of information retrieval systems [22][24]. However, relevance feedback for automatic summarization systems, which is a special application of these systems, has not been well studied. There are two types of relevance feedback models – the vector processing relevance feedback model, which makes use of term weights, and the probabilistic retrieval relevance feedback model, which uses purely probabilistic methods. We propose a relevance feedback model for fractal summarization that is based on these two existing relevance feedback models.

- As information is stored as vectors in most systems, the vector processing relevance feedback model is the most widely used. The system uses a query vector to specify relevant and irrelevant information, and relevance feedback is used to modify the query vector accumulatively, with documents being ranked subsequently according to their distance from the query vector. The best-known vector feedback algorithm is the Rocchio model [22], which measures similarity by using the inner product.
- Some researchers believe that documents should be extracted based on the theory of probability. Using this theory, documents are extracted based on the probability that each term will occur in relevant or irrelevant documents. This system uses relevance feedback to adjust the probability function of each term, and then recalculates the relevance probability of each document [24].

Relevance feedback models have been previously proposed for the four extraction features of fractal summarization that are discussed in detail in Section 2, but such a model that includes relevance feedback for the location, heading, and cue features is new to the field.

- The thematic feature displays a list of terms with thematic weight, which is equivalent to the query vector in the vector relevance feedback model. The summarization system extracts a set of sentences such that the inner product of the term list and the summary is maximized, and the list is constructed automatically. The weights of the terms are initialized as the *tfidf* score, which is adjusted accumulatively based on the relevance feedback to reflect the user's actual assignation of weight to the terms. The system increases the weights of the terms that appear in the sentences that have been selected, and decreases the weights of the terms that appear in the sentences that have been rejected [22][24]. The term list at the n+1$^{\text{th}}$ round (T^{n+1}) is constructed by the previous accumulated term list (T^n) and the accepted and rejected sentences in the n^{th} round feedback, i.e.,

  $$T^{n+1}{}_i = \alpha T^n{}_i + \beta P\left(t_i \middle| n^{th} \text{ round accepted}\right) - \gamma P\left(t_i \middle| n^{th} \text{ round rejected}\right),$$

  where α, β, γ are constants and $P(t_i)$ is the probability of term t_i. The thematic score of the sentence is the sum of *tfidf* score with relevance feedback, i.e., the score for the accumulated term list (T^n).
- In fractal summarization, documents are represented by a hierarchical tree structure. If many sentences under a given branch are accepted, then this branch is deemed to be more important, and its location score is increased. However, if many sentences under a given branch are rejected, then the location score of the branch is decreased. By using a conventional probabilistic relevance feedback model [4][24], the location score of a range block r is multiplied with a probability function $P(r)$ of range block r.

$$P(r) = \log\left(\frac{P(r|accepted)(1 - P(r|rejected))}{P(r|rejected)(1 - P(r|accepted))}\right),$$

where,

$$P(r|accepted) = \frac{no\ of\ sentences\ accepted\ in\ r + 0.5}{total\ no\ of\ sentences\ accepted + 1}, \text{ and}$$

$$P(r|rejecteded) = \frac{no\ of\ sentences\ rejected\ in\ r + 0.5}{total\ no\ of\ sentences\ rejected + 1}.$$

The relevance feedback on the location score is considered only for the calculation of the fractal value of a range block. When the system is extracting sentences inside a range block, it is disabled.

- The heading feature is an extension of the thematic feature to some extent because it represents a string of matching keywords in the corresponding heading. The heading weight of each term is the thematic weight with relevance feedback, which is propagated along the fractal structure, that is, the thematic weight of the term in the Fractal Sentence Heading Score is replaced with the weight of the accumulated term list that is constructed in the thematic feature.
- The cue feature has a dictionary with cue weights that have been defined by linguists. However, this may not truly reflect the user's preference, and therefore the cue weights are updated based on the sentences that are accepted or rejected in a similar manner as for the thematic feature.

5 Results

An experiment with ten participants was conducted to measure the performance of the summarization system with relevance feedback. The results show that relevance feedback greatly improves the precision of the summarization.

Usually, the performance of information systems is measured according to precision and recall. However, the performance of summarization systems is usually measured in

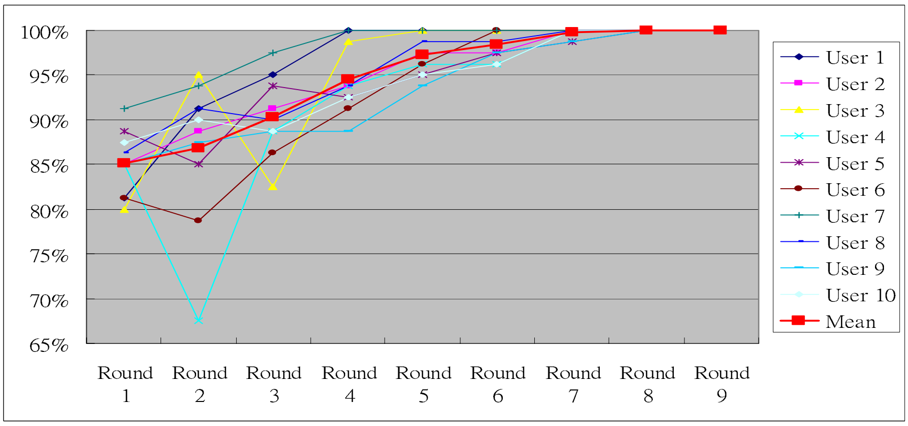

Fig. 2. Precision of Fractal Summarization with Relevance Feedback.

terms of precision only, as the measurement of recall is limited by the compression ratio of a summarization system. In addition to precision, the performance of relevance feedback can be measured by the time that is taken for the user to find information, and thus we measure the number of rounds that the summarization system takes to reach and retain its peak performance. First, a summary that was generated by generic fractal summarization was presented to the participants. The participants accepted or rejected the sentences based on whether they would include the sentence as part of the summary, and the precision of the summary was measured by the ratio of sentences that were accepted by the participants. The system then used the data of the users' feedback to update the sentence score, and generated another summary. This procedure was repeated until there was no further improvement. The results are shown in Figure 2.

As is shown in Figure 2, the average precision increases very quickly after the first few rounds. The mean precision in the first round is 85%, and increases significantly to 95% after three rounds of feedback. After that, the precision keeps increasing steadily, and reaches and retains 100% after the eighth round. By a *t*-test analysis of the precision, we found that there was no significant improvement in precision in the first and second rounds. The precision improved significantly at a 95% confidence level in the third to seventh round, and improved at an 80% confidence level in the eighth and ninth rounds. In other words, the performance is nearly saturated in the seventh round, after which the improvement in precision is not as significant as it in the previous rounds. In summary then, relevance feedback greatly improved the performance of the summarization system, and thus by using a fractal summarization model with relevance feedback, the professional abstractor can quickly extract important information from a document. This saves a lengthy read through the whole document, and allows the abstractor to generate a summary that is based on extracted sentences very quickly.

6 Conclusion

In this paper, a relevance feedback model for fractal summarization is proposed. Experiments were conducted with this model, and the results show that the relevance feedback model significantly improves the performance of fractal summarization. The employment of this model would make the automatic summarization system a much more useful tool for professional abstractors for the efficient generation of high quality abstracts that are based on extracted sentences.

References

[1] Barnsley M. F. and Jacquin, A. E. Application of Recurrent Iterated Function Systems to Images. Proc. SPIE Visual Comm. and Image Processing'88, 1001, 122-131, 1988.

[2] Baxendale P. Machine-Made Index for Technical Literature - An Experiment. IBM Journal (October), 354-361, 1958.

[3] Cowie J., Mahesh K., Nirenburg S., and Zajaz R. MINDS-Multilingual Interactive Document Summarization. Working Notes of the AAAI Spring Symposium on Intelligent Text Summarization. 131-132. California, USA, 1998.

[4] Cox D. et al. Analysis of Binary Data. 2nd Edition, Chapman & Hall, 1988.

[5] Craven T. C. Human Creation of Abstracts with Selected Computer Assistance Tools, Information Research, 3(4), 4, 1998.

[6] Craven T. C. Abstracts Produced Using Computer Assistance. J. of the American Soc. for Info. Sci., 51(8), 745-756, 2000.
[7] Edmundson H. P. New Method in Automatic Extraction. J. ACM, 16(2) 264-285, 1968.
[8] Endres-Niggemeyer B., Maier E., and Sigel A. How to Implement a Naturalistic Model of Abstracting: Four Core Working Steps of an Expert Abstractor. Information Processing and Management, 31(5) 631-674, 1995.
[9] Glaser B. G. and Strauss A. L. The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine de Gruyter, New York, 1967.
[10] Goldstein J. et al. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. Proc. SIGIR'99, 121-128, 1999.
[11] Hearst M. Subtopic Structuring for Full-Length Document Access. Proc. SIGIR'93, 56-68, 1993.
[12] Jacquin. A. E. Fractal Image Coding: a Review. Proc. IEEE, 81(10), 1451-1465, 1993.
[13] Kendall M., and Gibbons J.D. Rank Correlation Methods, 5th ed. New York: Edward Arnold, 1990.
[14] Koike, H. Fractal Views: A Fractal-Based Method for Controlling Information Display. ACM Tran. on Information Systems, ACM, 13(3), 305-323, 1995.
[15] Kupiec J. et al. A Trainable Document Summarizer. Proc. SIGIR'95, 68-73, Seattle, USA. 1995.
[16] Lam-Adesina M. and Jones G. J. F. Applying Summarization Techniques for Term Selection in Relevance Feedback. Proc. SIGIR 2001, 1-9, 2001.
[17] Lin Y. and Hovy E.H. Identifying Topics by Position. Proc. of Applied Natural Language Processing Conference (ANLP-97), Washington, DC, 283-290, 1997.
[18] Luhn H. P. The Automatic Creation of Literature Abstracts. IBM Journal of Research and Development, 159-165, 1958.
[19] Mandelbrot B. The Fractal Geometry of Nature. W.H. Freeman, New York, 1983.
[20] Morris G., Kasper G. M., and Adams D. A. The Effect and Limitation of Automated Text Condensing on Reading Comprehension Performance. Info. Sys. Research, 17-35, 1992.
[21] Ogden W., Cowie J., Davis M., Ludovik E., Molina-Salgado H., and Shin H. Getting Information from Documents You Cannot Read: an Interactive Cross-Language Text Retrieval and Summarization System. Joint ACM DL/SIGIR Workshop on Multilingual Information Discovery and Access, 1999.
[22] Rocchio J. Relevance Feedback in Information Retrieval. The Smart Retrieval System, 313-323, Prentice Hall, 1971.
[23] Salton G. and Buckley C. Term-Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, 24, 513-523, 1988.
[24] Salton G. et al. Improving Retrieval Performance by Relevance Feedback. J. America Soc. for Info. Sci., 41, 288-297, 1990.
[25] Teufel S. and Moens M. Sentence Extraction as a Classification Task. In Workshop of Intelligent and Scalable Text Summarization, ACL/EACL, 1997.
[26] Tsujimoto S. and Asada H. Understanding Multi-articled Documents, Proc. of the 10th Int. Conf. on Pattern Recognition, Atlantic City, N.J., 551-556, 1990.
[27] Wang F. L. and Yang C. C. Automatic Summarization of Chinese and English Parallel Documents, Proc. 6th Int. Conf. on Asian Digital Libraries, Kuala Lumpur, 2003.
[28] Yang C. C. and Li K. W. Automatic Construction of English/Chinese Parallel Corpora. J. of American Soc. for Info. Sci. and Tech., 54(8), 730-742, 2003.
[29] Yang C. C. and Wang F. L. Fractal Summarization for Mobile Device to Access Large Documents on the Web. Proc. 12th Int. WWW Conf., Budapest, Hungary, 2003.
[30] Yang, C. C. and Wang F. L. Fractal Summarization: Summarization Based on Fractal Theory, Proc. SIGIR 2003, Toronto, Canada, 2003.