# The Use of ART2 to Create Summaries from Texts of Different Areas

R. E. R. Christ, T. Bassani, J. C. Nievola, *Member IEEE*, C. N. Silla Jr.
*LASIN – Laboratório de Sistemas Inteligentes*
*PPGIA – Programa de Pós-Graduação em Informática Aplicada*
*PUCPR – Pontifícia Universidade Católica do Paraná*
*{rafaelchrist,tbassani,nievola,silla}@ppgia.pucpr.br*

## Abstract

*The volume of documents available electronically is growing fast, so it becomes difficult to access and select desired information in a fast and efficient way. In this context the automatic summarization task assumes a very imperative role; therefore one seeks to reduce the size of a document, preserving to the maximum its informative content. In this paper, it's applied a model which uses sentence clusters from an ART2 neural network to generate extractive summaries. Different models can be developed from distinct area documents. Hence, the aim of this work is to evaluate the performance of those models when they summarize documents from correlated or non correlated areas.*

## 1. Introduction

Once the volume of documents available electronically is growing fast, it becomes extremely important to access and select desired information in a fast and efficient way. Thus, automatic document summarization plays a key role in this new context.

Summarization—the art of abstracting key content from one or more information sources—has become an integral part of everyday life. People keep abreast of world affairs by listening to news. They base investment decisions on stock market updates. They even go to movies largely on the basis of reviews they've read. With summaries, they can make effective decisions in less time [7].

At the most basic level, summaries differ according to whether they are extracts or abstracts. Both kinds of summarization have two core tasks: determine what is salient in the source being summarized and decide how to reduce its content.

The abstract approach, the creation of a summary directly from documents demands a deep knowledge of natural language and is still beyond the competency of actual computer systems [8]. We choose the extractive approach for generating summaries rather than the abstract approach because the first one is viable and simpler [9]. Nevertheless, most of the systems projected to summarize documents tries to generate an ideal extractive summary from a specific document by using statistical techniques and superficial linguistic analysis [6].

To address these limitations, researchers are looking at a variety of approaches, which roughly fall into two categories. One is the knowledge-poor approaches, which rely on not having to add new rules for each new application domain or language. The second one, the knowledge-rich approaches, assumes that if you grasp the meaning of the text, you can reduce it more effectively, thus yielding a better summary. Both approaches rely on a sizeable knowledge base of rules, which must be acquired, maintained, and then adapted to new applications and languages [7]. Nevertheless, we use a rather simpler and effective architecture to generate extractive summaries.

An ART2 neural network is used to create sentence clusters which are supposed to have similar syntactic and semantic features. These clusters will be employed to generate an extractive summary from a given document. This model is implemented by training the neural network with a base document. Once the ART2 neural network is trained, the document corresponding to the desired summary can be presented to the neural network.

Generally, it's assumed that the summary quality depends on the initial base. Most of the times, the base and the summarized documents are from the same area. However, no work known by these authors analyzed whether or not this presupposition is valid. Consequently, the aim of this work is to evaluate the performance of different models when they summarize documents from correlated or non correlated areas.

## 2. Model Architecture

The Figure 1. shows how the model architecture woks.

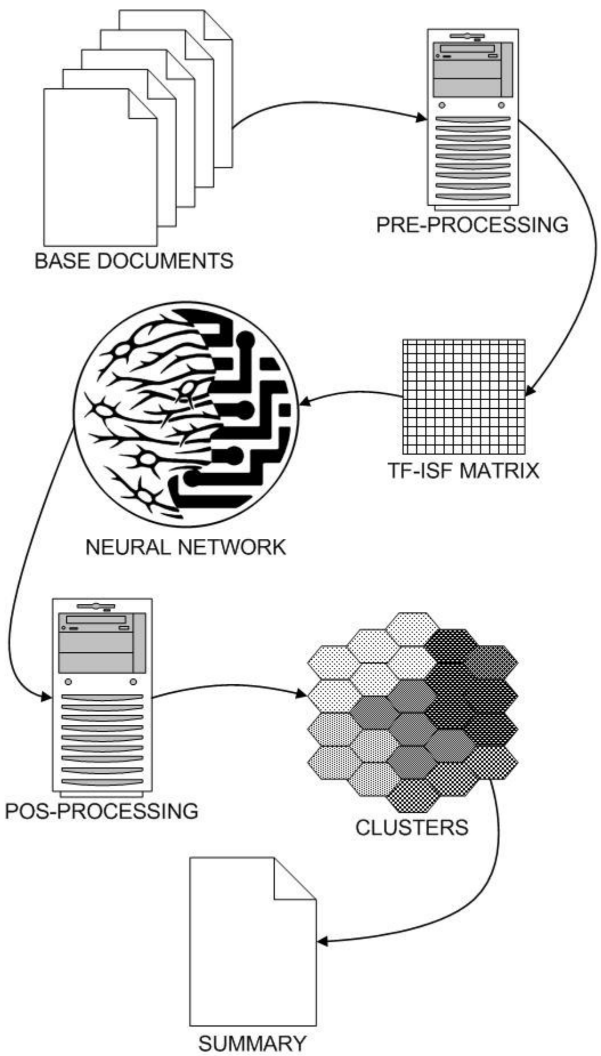


**Figure 1. Model Architecture**

The main advantage of automatic text summarization is avoiding interference and subjectivity from a human judge and allowing one to easily create summaries of many different sizes. The summarization task may be divided in 3 stages [6]:

1. Creation of a representation from the original document.
2. Conversion of this representation to a summary representation.
3. Conception of a summary from the summary representation.

Following these 3 steps, one may be able to build an extractive summary. In this work, the 2 first steps are assembled together. The design of a representation from the original document and its conversion to a summary version involves a measure named TF-ISF which is derived from the TF-IDF (Term Frequency - Inverse Term Frequency). The TF-IDF is generally utilized in information retrieval systems [10]. From now on, the TF-ISF from a word w in a sentence s is defined in (1) as:

$$TFISF(w,s) = TF(w,s) \times ISF(w) \qquad (1)$$

TF(w,s) is the frequency of occurrences of the word w in the sentence s and ISF(w) is described in (2):

$$ISF(w) = \log\left(\frac{|s|}{SF(w)}\right) \qquad (2)$$

Finally, SF(w) is the number of sentences where the word w occurs.

### 2.1. Pre-processing

The pre-processing is employed in order to transform a document into an ART2 network input. It also takes part in a rather essential role in the quality of the summary as will be further seen. The pre-processing may be split in 4 distinct steps:

1. Stop-words removal
   Stop-words are words that, even though show up frequently in a text, have no concrete significance. E.g. “what”, “to” and “they”. Since these words are useless for the summary construction, they must be taken away from its creation. The algorithm employed makes use of more than 200 stop-words.
2. Stemming algorithm
   The stemming algorithm removes any prefixes and suffixes from the text words. After its execution, all text words are converted to their radical form. We utilized the Porter algorithm which is a known stemming algorithm in the literature [11].
3. Case-fold removal
   The case-fold removal algorithm converts all the document words to a customary format. It means that, for example, the words Theory, theory and THEORY will be replaced by the word theory.
4. Vector of words and vector of TFs assembling
   The vector of words contains all the distinct words from the base document that will be used to train the ART2 neural network. Each vector of words has its own vector of TF (Term

Frequencies). It contains the frequency of occurrence of the words in the base document.

5. m x n TF-ISF matrix building
   In the TF-ISF matrix, m is the number of sentences of a document and n is the length of the vector of words from the base document. Hence, an ART2 input is a sentence versus words 2D matrix. A line in this matrix corresponds to a sentence in a document. It contains the TF-ISF from each single word in a given sentence.

### 2.2. ART2 neural network

ART2 is an unsupervised neural network algorithm derived form the resonance theory [1]. The ART networks are rather good at pattern recognizing and pattern classification. Their design allows the user to control the similarity between the patterns accepted by the same cluster [2]. ART2 can learn about significant new classes, yet remain stable in response to previously learned classes [3]. Thus, it's able to meet the challenges of text summarization where numerous variations are common.

ART networks are configured to recognize invariant properties of a given problem domain; when presented with data pertinent to the domain, the network can categorize it on the basis of this features. This process also categorizes when distinctly different data are presented and it includes the ability to create new clusters. ART networks accommodate these requirements through interactions between different subsystems, designed to process previously encountered and unfamiliar events, respectively [4].

We choose the ART2 neural network rather than other classifiers once it's capable of incrementally increasing the numbers of clusters if needed [2] [5]. This is a rather important feature when the size of the input vectors can vary a lot. It happens here since the length of the documents employed in the summarization process may be very different.

ART2 networks were designed to process continuous input pattern data. A special characteristic of such networks is the plasticity that allows the system to learn new concepts and at the same time retain the stability that prevents destruction of previously learned information.

### 2.3. Pos-processing

The pos-processing step is where actually the summaries are generated. After the ART2 neural network is trained, a candidate document to have the summary generated can be presented. The ART2 neural network clusters the sentences from this document and also provides their output activation.

This activation measures the response of the neural network's cluster to a given sentence. For every sentence its activation and the cluster where it is located are known. The pos-processing algorithm is carried out by:

1. Sorting the clusters' sentences from the highest output activation to the lowest output activation.
2. Sorting all the clusters based on its first sentence's output activation.
3. Creating a summary from the sentences selected previously. The summary size is determined by the user.

## 3. Experiments

### 3.1. Document Bases

For the experiments it were assembled 3 distinct document bases, referred as A, B and C. The base A and B are specialist bases. It means that they only hold texts from the same area. However the base C is more generic because it is a mixed text base with texts from the base A and B. A description of each base is available as follow:

A. Document base A
   The base A attaches medicine articles about myocardial therapies and diseases. A specialist from this area selected 23 articles from [15]. Each one contains abstracts with an average size of 288 words and the main text has 2811 singular words distributed in 2981 sentences.
B. Document base B
   The base A contains 27 technical articles about computer networking selected from the Ziff database [12]. These articles have nearly 550 words and for each one there is a summary produced by its author. It has 2740 singular words distributed in 1985 sentences.
C. Document base C
   The base C was formed by 12 articles from the base A plus 12 articles from the base B. It has 4555 singular words distributed in 3383 sentences.

We also selected 2 texts that will be used to evaluate the performance of the 3 models in our experiments. The first document, referred as Text 1, was extract from the base A. The size of this document is 190 sentences. The second document, Text 2, was extract from the base B and has 222 sentences.

One might notice that these 2 texts are no longer part of the base A and B. It means that they are not employed in the ART2 neural network's training. Thus, they are only presented to the neural network in the validation step.

## 3.1. Description

The experiments are divided in 2 steps, the first step is the training (after the database pre-processing) and the second step is the validation. The first step is implemented by training the ART2 neural network with a base document. Then, once the neural network is trained, we move to the second step. In the validation step, singular documents can be presented to the network to have their summaries generated.

In the first step the ART2 neural network was executed 50 epochs in slow learning mode. The parameters of the neural network were kept the same through the tests, except for the vigilance parameter ($\rho$).

The vigilance parameter is the main parameter of the ART2 neural network. It determines how many clusters will be formed [2]. Thus, it has a rather important role in the performance of the neural network. In order to avoid the interference of this parameter in the texts, 7 different vigilance parameters were utilized, 0.80, 0.84, 0.91, 0.94, 0.97 and 0.99.

We implemented 3 models with the bases A, B and C. In each model the neural network was trained with a distinct $\rho$. After that, we utilized the documents Text 1 and Text 2 in the validation step.

# 4. Results & Discussion

In order to avoid possible subjectivity and interference of a human judge in the evaluation of each generated summary, we utilized automatically ideal extractive summaries using the authors' summaries. With that, it is possible to apply the Precision and Recall metrics to evaluate each generated summary, which is often used in other experiments in the literature like in [13] and [14].

Let NCS be the number of correctly selected sentences in the produced summary; be NS the number of selected sentences in the produced summary; and be NSI the number of sentences in the ideal extractive summary. Hence, Precision and Recall are defined in (3) and (4) as:

$$precision = \frac{NCS}{NS} \qquad (3)$$

$$recall = \frac{NCS}{NSI} \qquad (4)$$

In our experiments, since the automatic summaries and the produced summaries have the same compression rate (of 10%), Precision is equal to Recall.

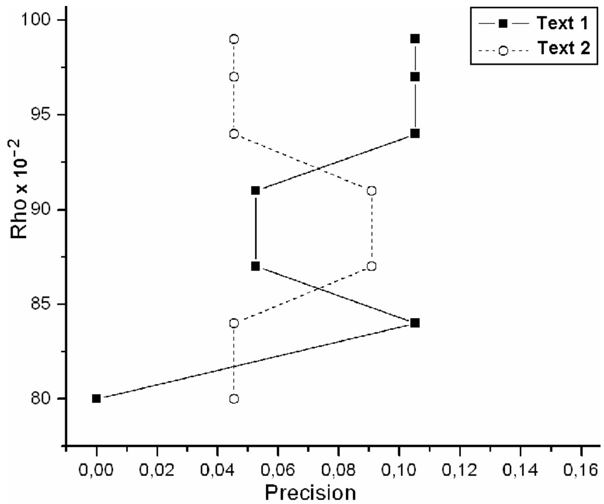

**Figure 2. Precision *versus* ρ for Base A**

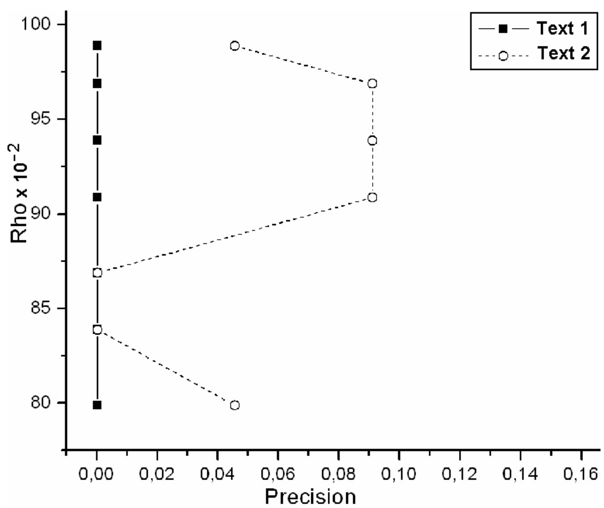

**Figure 3. Precision *versus* ρ for Base B**

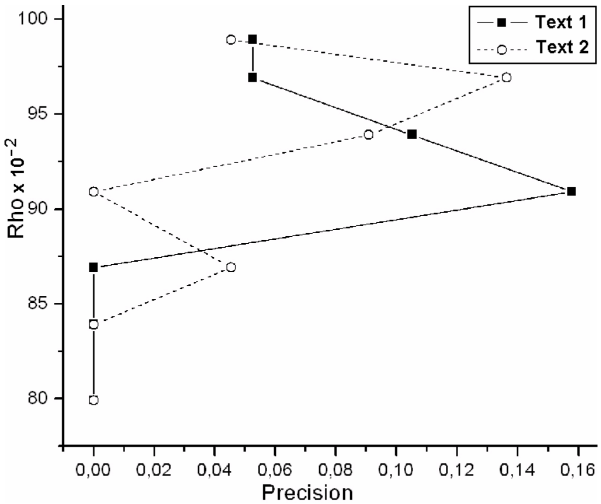


**Figure 4. Precision *versus* ρ for Base C**

The relation between Precision and ρ showed that low values of ρ (here something below 0.85) corresponds to situations where the results are tragic, corresponding to summaries with very low precision. The results with ρ high (in this case above 0.99) caused also troubles in the results. As is pointed in the literature, there is a narrow range of ρ that gives the better results, regarding the clustering of the input elements. In our case, this clustering had a direct influence in the summary, and its slight variation causes different results.

An interesting aspect of the obtained results is the fact that the best results both are from the mixed texts base. This is an interesting aspect from the summarization point of view, which would allow having a trained ART2 summarizer in different subjects that would produce good extractive summaries.

When the system was trained with base A (myocardial infection base) the better result corresponded to the summary of the myocardial infection article. The same thing happened when the system was trained with base B (Ziff base), that means, the better summary was the one derived from the Ziff base. Both results agree with what was expected.

Despite that, the use of a mixed base in the training showed something interesting. The generated summaries got better precision than in the first two cases (which were approximately equal). This finding could be interpreted as indicating that one should utilize a base that has documents from a multitude of domains when generating a specific summary. This is true even when the desired summary comes from documents in a narrow domain.

In the model implemented with the base A, the neural network put the sentences of Text 1 in an average of 72 distinct clusters. Also, it was not able to cluster 18% of the sentences. However, for the Text 2, it put the sentences in only 17 distinct clusters and it was no able to cluster 52% of the sentences.

The same behavior was observed in the model implemented with the base B. The document of the same area of the model was clustered in a bigger number of singular clusters (62) than the document from a non correlated area (13). Also, the text of a different area had most of its sentences rejected by the neural network.

For the mixed text base, the Text 1 and Text 2 presented a similar behavior to the cases when they were from the same area of the model. The most relevant fact to notice is that the clusters created here have sentences from the both bases. This is illustrated in the Figure 5.

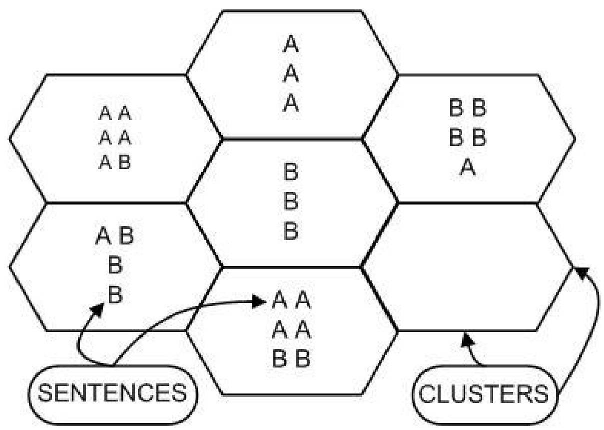


**Figure 5. Clustering of the sentences in the mixed text model neural network**

The neural network of the mixed text base is more plastic than the network of the specialist bases. Along its training, the plasticity of the neural network allowed it to create clusters that still are well defined: most of the sentences in one cluster are from the same area, but these clusters are more generic. So, they will be more likely to receive a new sentence than the clusters from a neural network trained with a specialist base.

## 5. Conclusion

The first point to note is that this work is not trying to create the most efficient summarizer, as is the case most of the times in the information retrieval literature.

Instead, the goal was to verify whether there is a relation between the quality of the generated summary and the specificity of the domain used to create the model.

In order to do so, some experiments were made. They showed that focused training documents create models which originate summaries biased towards that domain. It implies that the use of such models is not reliable when applied in documents from another domain.

The experiments also showed that a training base with documents derived from more than one area could create a model better suited to the generation of summaries for these areas, and that the quality of them is even better than that of the first case.

For these results to be considered general more bases should be used and other kinds of summarizers employed.

## 6. References


[1] Carpenter, G. A., Grossberg, S., "ART2: Self-organization of Stable Category Recognition Codes for Analog Input Patterns", Applied Optics, 1987b.

[2] Fausett, L.V., Fundamentals of Neural Networks Architectures, Algorithms, and Applications. New Jersey: Prentice Hall International Inc, New Jersey, p.246-287, 1994.

[3] Chang, H., Kopaska-Merkel, D. C., Chen, H., Durrans, R. D., "Lithofacies Identification Using Multiple Adaptative Resonance Theory Neural Networks and Group Decision Expert System", Computers & Geocites 26, 2000.

[4] Avazdavani, A., Mortazavi, S. S., "Application of Modified Art2 Artificial Neural Network In Classification of Structural Members", 15th ASCE Engineering Mechanics Conference, Columbia University, New York, NY, June 2-5, 2002.

[5] Bingus, J.P.; 1996. Data Mining with neural networks: Solving Business Problems - From an aplication development decision support. McGraw-Hill Book Co, New York, p.220.

[6] Jones, K. S., "Automatic Summarizing: factor and directions", In I. Mani e M. Baybury editors, Advances in Automatic Text Summarization, MIT Press, 1999.

[7] Hahn, U., Mani, I., "The Challenges of Automatic Summarization", Computer, IEEE Press, Volume: 33, Issue: 11, p.29-36, Nov. 2000.

[8] Mitra, M.; Singhal, A.; Buckley, C., “Automatic text summarization by paragraph extraction”, In Proceedings of the ACL’97/EACL’97 Workshop on Intelligent Scalable Text Summarization, 1997.

[9] Edmunson, H. P., New methods in automatic extracting, Journal of the Association for Computing Machinery 16, p. 264-285, 1969.

[10] Neto, J. L., “Contribuição no Estudo de Técnicas para Sumarização Automática de Textos”, Dissertação de Mestrado, Programa de Pós Graduação em Informática Aplicada, PUC-PR, 2002.

[11] Porter, M., (2000). An Algorithm for Suffix Stripping. [Online] Available: http://telemat.det.unifi.it/book/2001/wchange/download/stem_porter.html

[12] Ziff Davis document base information, (2002). [Online] Available: http://www.ziffdavis.com

[13] Larocca Neto, J.; Freitas, A. A.; Kaestner, C. A. A., “Automatic Text Summarization using a ML Approach”, In the Proc. of the XVI Brazilian Symposium on Artificial Intelligence, Lecture Notes on Compute Science, No. 2507, pp. 205-215, 2002.

[14] Radev, D. R., at al., “Evaluation Challenges in Large-scale Document Summarization”, In the Proc. of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 375-382, July 2003.

[15] The Lancet journal, (2004). [Online] Available: [http://www.thelancet.com].