Experimenting Text Summarization on Multimodal Aggregation

Giuliano Armano, Alessandro Giuliani, Alberto Messina, Maurizio Montagnuolo and Eloisa Vargiu

Abstract Nowadays, Web is characterized by a growing availability of multimedia data together with a strong need for integrating different media and modalities of interaction. Hence, the main goal is to bring into the Web data thought and produced for different media, such as TV or radio content. In this scenario, we focus on multimodal news aggregation retrieval and fusion. In particular, we present preliminary experiments aimed at automatically suggesting keywords to news and news aggregations. The proposed solution is based on the adoption of extraction-based text summarization techniques. Experiments are aimed at comparing the selected text summarization techniques with respect to a simple technique based on part-of-speech tagging. Results show that the proposed solution performs better than the baseline solution in terms of precision, recall, and F1.

1 Introduction

Modern broadcasters are facing an unprecedented technological revolution from traditional dedicated equipment to commodity hardware and software components, and from yesterday-one-to-many delivery paradigms to nowadays-Internet-based interactive platforms. In this challenging scenario, information engineering and integration plays a vital role in optimizing costs and quality of the provided services, and in reducing the "time to market" of data.

In this scenario, this paper focuses on multimodal news aggregation, retrieval, and fruition. Multimodality is intended as the capability of processing, gathering,

G. Armano, A. Giuliani and E. Vargiu
University of Cagliari, Dept.of Electrical and Electronic Engineering, Piazza d'Armi, I09123 Cagliari (Italy) e-mail: {armano, alessandro.giuliani, vargiu}@diee.unica.it

A. Messina and M. Montagnuolo
RAI Centre for Research and Technological Innovation, C.so Giambone, 68, I10135 Torino (Italy) e-mail: {a.messina, maurizio.montagnuolo}@rai.it

manipulating, and organizing data from multiple media (e.g., television, radio, the Internet) and made of different modalities such as audio, speech, text, image, and video. In particular, we present a preliminary study aimed at automatically generating tag clouds for representing the content of multimodal aggregations (MMAs) of news information from television and from the Internet. To this end, we propose a solution based on Text Summarization (TS) and we make experiments to compare classical extraction-based TS techniques with respect to a simple technique based on part-of-speech (POS) tagging.

The rest of the paper is organized as follows. Section 2 recalls relevant work on multimedia semantics, information fusion, heterogeneous data clustering, and text summarization. In Section 3, we recall the model for multimodal aggregation previously presented in [26] and we illustrate how news are stored according to that model. Section 4 focuses on the problem addressed in this paper by describing the adopted extraction-based TS techniques. In Section 5, we illustrate our experiments aimed at exploiting TS in MMA. Section 6 ends the paper reporting conclusions and future research directions.

2 Background

2.1 Multimedia Semantics

Recently, several research activities have attempted to provide the state-of-the-art of content-analysis-based extraction of multimedia semantics, with the stated intention to provide a unified perspective to the field [10, 12, 19, 31]. Mostly, these works succeed in giving a complete and updated panorama of the existing techniques based on content analysis for multimedia knowledge representation. In our opinion, the work done so far has only partially achieved the objective of giving a deep understanding of problems related to multimedia semantics. This statement comes from the observation that only a very few research solutions and tools end up to be useful for practical purposes in the media industry. In our opinion, this is due to a significant lack of precision in the definition of relevant problems, which led to huge research efforts, but only seldom in directions exploitable by the media industry (e.g., broadcasters, publishers, producers) in a straightforward way. Emergent technologies like Omni-Directional Video [13] enhance the urgency of a high-level re-elaboration of the discipline.

Modern research efforts in multimedia information retrieval (MIR) have been recently summarized by Lew et al. [19]. One of the key issues pointed out by the authors is the lack of a common and accepted test set for researchers conducting experiments in the field of MIR. The somewhat central claim of Lew et al. is that published test sets are typically scarcely relevant for real-world applications, so that this situation may bring in the risk to see the research community around MIR to be "*isolated from real-world interests*" in the near future. This claim sounds as a

serious alarm bell for researchers and practitioners of the field. Let us also consider that in concrete scenarios, as the one proposed in [24], the accuracy figures obtained by state-of-the-art tools may not be fully satisfactory for an industrial exploitation [17, 23]. Lew et al. give also an interesting hint on some future research directions, including human centered methods, multimedia collaboration, neuroscience methods exploitations, folksonomies.

Integration between semantic Web technologies and multimedia retrieval techniques is considered a future challenge by many researchers [32]. In this field, the task of concept detection is concerned with identifying of instances of semantically evocative language terms through the numerical analysis of multimedia items. The work of Bertini et al. [5] proposes a solution in the domain of television sport programmes. Their approach uses a static hierarchy of classes (named pictorially enriched ontology) to describe the prototypical situations findable in football matches and associate them with low-level visual descriptors. In [6], the authors present a complete system for creating multimedia ontologies, automatic annotation, and video sequences retrieval based on ontology reasoning.

2.2 Information Fusion and Heterogeneous Data Clustering

Information (or data) fusion can be defined as the set of methods that combine data from multiple sources and use the obtained information to discover additional knowledge, potentially not discoverable by the analysis of the individual sources.

First attempts to organize a theory have been done in [29], in which the author proposes a cross-document structure theory and a taxonomy of cross-document relationships. Recently, some proposals have been made to provide a unifying view. The work in [14] classifies information fusion systems in terms of the underlying theory and formal languages. Moreover, in [22], the author describes a method (Finite Set Statistics) which unifies most of the research on information fusion under a Bayesian paradigm.

Many information fusion approaches currently exist in many areas of research, e.g., multi-sensor information fusion, notably related to military and security applications, and multimedia information fusion. In the latter branch, the closest to the present research, the work in [33] analyses best practices for selection and optimization of multimodal features for semantic information extraction from multimedia data. More recent relevant works are [28] and [18]. In [28], the authors present a self-organizing network model for the fusion of multimedia information. In [18], the authors implement and evaluate a fusion platform implementing the framework within a recommendation system for smart television in which TV programme descriptions coming from different sources of information are fused.

Heterogeneous data clustering is the usage of techniques and methods to aggregate data objects that are different in nature, for example video clips and textual documents. A type of heterogeneous data clustering is co-clustering, which allows simultaneous clustering of the rows and columns of a matrix. Given a set of m rows

in n columns, a co-clustering algorithm generates co-clusters, i.e., a subset of rows which exhibit similar behavior across a subset of columns, or vice-versa. One of the first methods conceived to solve the co-clustering of documents using word sets as features is represented by [20], where RSS items are aggregated according to a taxonomy of topics. More challenging approaches are those employing both cross-modal information channels, such as radio, TV, the Internet, and multimedia data [9, 34].

2.3 Text Summarization

Radev et al. [30] define a summary as "a text that is produced from one or more texts, that conveys important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that". This simple definition highlights three important aspects that characterize the research on automatic summarization: (i) summaries may be produced from a single document or multiple documents; (ii) summaries should preserve important information; and (iii) summaries should be short. Unfortunately, attempts to provide a more elaborate definition for this task are in disagreement within the community [8].

Summarization techniques can be divided in two groups [16]: those that extract information from the source documents (*extraction-based approaches*) and those that abstract from the source documents (*abstraction-based approaches*). The former impose the constraint that a summary uses only components extracted from the source document. These approaches put strong emphasis on the form, aiming to produce a grammatical summary, which usually requires advanced language generation techniques. The latter latter relax the constraints on how the summary is created. These approaches are mainly concerned with what the summary content should be, usually relying solely on extraction of sentences.

Although potentially more powerful, abstraction-based approaches have been far less popular than their extraction-based counterparts, mainly because generating the latter is easier. While focusing on information retrieval, one can also consider topic-driven summarization, which assumes that the summary content depends on the preferences of the user and can be assessed via a query, making the final summary focused on a particular topic. Since in this paper we are interested in extracting suitable keywords, we exclusively focus on extraction-based methods.

An extraction-based summary consists of a subset of words from the original document and its bag of words (BoW) representation can be created by selectively removing a number of features from the original term set. Typically, an extraction-based summary whose length is only 10-15% of the original is likely to lead to a significant feature reduction as well. Many studies suggest that even simple summaries are quite effective in carrying over the relevant information about a document. From a text categorization perspective, their advantage over specialized feature selection methods lies in their reliance on a single document (the one that is being summarized) without computing the statistics for all documents sharing the same category

label, or even for all documents in a collection. Moreover, various forms of summaries become ubiquitous on the Web and in certain cases their accessibility may grow faster than that of full documents.

Earliest instances of research on summarization of scientific documents extract salient sentences from text using features like word and phrase frequency [21], position in the text [4], and key phrases [11]. Various works published since then had concentrated on other domains, mostly on newswire data [27]. Many approaches addressed the problem by building systems dependent on the type of the required summary.

3 Multimodal Aggregation

Multimodal aggregation of heterogeneous data, also known as *information mash-up*, is a hot topic in the World Wide Web community. A multimodal aggregator is a system that merges content from different data sources (e.g., Web portals, IPTV, etc.) to produce new, hybrid data that was not originally provided. Here, the challenge lies in the ability of combining and presenting heterogeneous data coming from multiple information sources, i.e., *multimedia*, and consisting of multiple types of content, i.e., *cross-modal*. As a result of this technological breakthrough, the content of modern Web is characterized by an impressive growth of multimedia data, together with a strong trend towards integration of different media and modalities of interaction. The mainstream paradigm consists in *bringing into the Web* what was thought (and produced) for different media, like TV content (acquired and published on websites and then made available for indexing, tagging, and browsing). This gives rise to the so-called *Web Sink Effect*. This effect has rapidly started, recently, to unleash an ineluctable evolution from the original concept of the Web as a resource where to *publish* things produced in various forms *outside* the Web, to a world where things *are born and live* on the Web. In this paper, we adopt Web newspaper articles and TV newscasts as information sources to produce multimodal aggregations of informative content integrating items coming from both contributions. In the following of this section we briefly overview the main ideas behind this task, and we point the interested reader to [26] for details.

The corresponding system can be thought as a processing machine having two inputs, i.e., digitized broadcast news streams (DTV) and online newspapers feeds (RSSF), and one output, i.e., the multimodal aggregations that are automatically determined from the semantic aggregation of the input streams by applying a co-clustering algorithm whose kernel is an asymmetric relevance function between information items [1].

Television news items are automatically extracted from the daily programming of several national TV channels. The digital television stream is acquired and partitioned into single programmes. On such programmes, newscast detection and segmentation into elementary news stories is performed. The audio track of each story

is finally transcribed by a speech-to-text engine and indexed for storage and retrieval. Further details can be found in [25].

The RSSF stream consists of RSS feeds from several major online newspapers and press agencies. Each published article is downloaded, analyzed, and indexed for search purposes. The first step of the procedure consists in cleaning the downloaded article Web pages from boilerplate content, i.e., HTML markups, links, scripts, and styles. Linguistic analysis, i.e., sentence boundary detection, sentence tokenization, word lemmatization and POS tagging, is then performed on the extrapolated contents. The output of this analysis is then used to transform the RSS content into a query to access the audio transcriptions of the DTV news stories, thus allowing to combine text and multimedia in an easy way.

The output of the clustering process is a set of multimodal aggregations of broadcast news stories and newspaper articles related to the same topic. TV news stories and Web newspaper articles are fully cross-referenced and indexed. For each multimodal aggregation, users can use automatically extracted tag clouds, to perform local or Web searches. Local searches can be performed either on the specific aggregation the tags belong to or to the global set of discovered multimodal aggregations. Tag clouds are automatically extracted from each thread topic as follows: (i) each word classified as proper noun by the linguistic analysis is a tag; (ii) a tag belongs to a multimodal aggregation if it is present in at least one aggregated news article; and (iii) the size of a tag is proportional to the cumulative duration of television news items which are semantically relevant to the aggregated news article to which the tag belongs. In so doing, each news aggregation, also called *subject*, is described by a set of attributes, the main being:

- *info*, the general information included title and description;
- *categories*, the set of most relevant categories to which the news aggregation belong. They are automatically assigned by AI:Categorizer[1], trained with radio programme transcriptions, according to a set of journalistic categories (e.g., *Politics*, *Currents Affairs*, *Sports*);
- *tagclouds*, a set of automatically-generated keywords;
- *items*, the set of Web articles that compose the aggregation;
- *videonews*, the collection of relevant newscasts stories that compose the news aggregation.

Hence, a news aggregation is composed by online articles (*items*) and parts of newscasts (*videonews*). In this paper, we concentrate only in the former. Each item is described as set of attributes, such as:

- *pubdate*, the timestamp of first publication;
- *lastupdate*, the timestamp when the item was updated;
- *link*, the URL of the news Web page;
- *feed*, the RSS feed link that includes the item;
- *title*, the title;
- *description*, the content;

[1] http://search.cpan.org/ kwilliams/AI-Categorizer-0.09/lib/AI/Categorizer.pm

- *category*, the category to which the news belong (according to the previously mentioned classification procedure);
- *keywords*, the keywords automatically extracted as described above.

4 Text Summarization in Multimodal Aggregation

In this paper, we are interested in automatically suggesting keywords to news and news aggregations in the area of news distribution and retrieval. In particular, we are aimed at selecting keywords relevant to the news and news aggregations. Among other solutions, we decided to use suitable extraction-based TS techniques. To this end, we first consider six straightforward but effective extraction-based text summarization techniques proposed and compared in [16] (in all cases, a word occurring at least three times in the body of a document is a keyword, while a word occurring at least once in the title of a document is a title-word):

- *Title* (T), the title of a document;
- *First Paragraph* (FP), the first paragraph of a document;
- *First Two Paragraphs* (F2P), the first two paragraphs of a document;
- *First and Last Paragraphs* (FLP), the first and the last paragraphs of a document;
- *Paragraph with most keywords* (MK), the paragraph that has the highest number of keywords;
- *Paragraph with Most Title-words* (MT), the paragraph that has the highest number of title-words.

Let us note that we decided to not consider the Best Sentence technique, i.e. the technique that takes into account sentences in the document that contain at least 3 title-words and at least 4 keywords. This method was defined to extract summaries from textual documents such as articles, scientific papers and books. In fact, news are often inadequate to find meaningful sentences composed by at least 3 title-words and 4 keywords in the same sentence.

Furthermore, we consider the enriched techniques proposed in [2]:

- *Title and First Paragraph* (TFP), the title of a document and its first paragraph:
- *Title and First Two Paragraphs* (TF2P), the title of a document and its first two paragraphs;
- *Title, First and Last Paragraphs* (TFLP), the title of a document and its first and last paragraphs;
- *Most Title-words and Keywords* (MTK), the paragraph with the highest number of title-words and that with the highest number of keywords.

One may argue that the above methods are too simple. However, as shown in [7], extraction-based summaries of news articles can be more informative than those resulting from more complex approaches. Also, headline-based article descriptors proved to be effective in determining user's interests [15]. Moreover, these techniques have been successfully applied in the contextual advertising field [3].

5 Experiments and Results

To assess the effectiveness of TS in the task of suggesting relevant keywords to news and news aggregations, we perform some comparative experiments. In particular, we performed two sets of experiments: (i) experiments on the sole news comparing the performance with those corresponding to the adoption of the keywords provided in the *keyword* attribute and (ii) experiments on news aggregations comparing the performance with those corresponding to the adoption of the keywords provided in the *tagclouds* attribute. Results have been calculated in terms of precision, recall, and F1 by exploiting a suitable classifier.

Experiments have been performed on about 45,000 Italian news and 4,800 news aggregations from January 16, 2011 to May 26, 2011. The adopted dataset is composed by XML files, each one describing a subject according to the attributes described in Section 3. News and news aggregations were previously classified into 15 categories, i.e., the same categories adopted for describing news and news aggregations.

5.1 Experimenting Text Summarization on News

Experiments on news have been performed by adopting a system that takes as input an XML file that contains all the information regarding a news aggregation. For each TS technique, first the system extracts the news, parses each of them, and adopts stop-word removing and stemming. Then, it applies the selected TS technique to extract the corresponding keywords in a vector representation (*BoW*). To calculate the effectiveness of that technique, the extracted *BoW* is given as input to a centroid-based classifier, which represents each category with a centroid calculated starting from a suitable training set[2]. A *BoW* vector is then classified by measuring the distance between it and each centroid, by adopting the cosine similarity measure.

Performances are calculated in terms of precision, recall, and F1. As for the baseline technique (B), we considered the BoW corresponding to the set of keywords of the *keywords* attribute. Table 1 summarizes the results.

Table 1 Comparisons among TS techniques on news.

	B	**T**	**FP**	**F2P**	**FLP**	**MK**	**MT**	**TFP**	**TF2P**	**TFLP**	**MTK**
P	0.485	0.545	0.625	0.705	0.681	0.669	0.650	0.679	**0.717**	0.706	0.692
R	0.478	0.541	0.594	0.693	0.667	0.665	0.640	0.663	**0.704**	0.697	0.686
F1	0.481	0.543	0.609	0.699	0.674	0.667	0.645	0.671	**0.710**	0.701	0.689
# terms	5	5	13	23	22	20	15	16	25	24	18

[2] In order to evaluate the effectiveness of the classifier, we performed a preliminary experiment in which news are classified without using TS. The classifier shown a precision of 0.862 and a recall of 0.858.

5.2 Experimenting Text Summarization on News Aggregations

Experiments on news aggregations have been performed in a way similar to the one adopted for the sole news. For each TS technique, first the system processes each news belonging to the news aggregation in order to parse it, to disregard stop-words, and to stem each remaining term. Then, it applies to each news the selected TS technique in order to extract the corresponding keywords in a *BoW* representation. Each extracted *BoW* is then given in input to the same centroid-based classifier used for the news. The category to which the news aggregation belongs to is then calculated averaging the scores given by the classifier for each involved item.

Table 2 shows the results obtained by comparing each TS technique, the baseline (B) being the BoW corresponding to the set of keywords of the *tagclouds* attribute.

Table 2 Comparisons among TS techniques on news aggregations.

	B	**T**	**FP**	**F2P**	**FLP**	**MK**	**MT**	**TFP**	**TF2P**	**TFLP**	**MTK**
P	0.624	0.681	0.693	0.764	0.734	0.717	0.727	0.731	**0.770**	0.766	0.737
R	0.587	0.678	0.683	0.766	0.728	0.709	0.718	0.728	**0.769**	0.759	0.729
F1	0.605	0.679	0.688	0.765	0.731	0.713	0.723	0.729	**0.769**	0.762	0.733
# terms	62	70	204	319	337	231	200	215	318	338	280

5.3 Discussion

Results clearly show that, for both news and news aggregations, TS improves performances with respect to the adoption of the baseline keywords. In particular, best performances in terms of precision, recall, and –hence– F1, are obtained by adopting the TF2P technique. The last row of Table 1 and Table 2 shows the number of terms extracted by each TS technique. It is easy to note that, except for the T technique, TS techniques extract a number of terms greater than that extracted by the baseline approach. Let us also note that precision, recall, and F1 calculated for news aggregations are always better than those calculated for news. This is due to the fact that news aggregations are more informative than the sole news and the number of extracted keywords is greater.

To better highlight the adopted extraction techniques, Figure 1 shows the description of a news aggregation[3], a selection of its tag clouds, and a selection of the keywords extracted by the most effective TS technique, i.e., TF2P.

[3] Actually, we are using Italian news but, for the sake of clarity, we decided to translate it

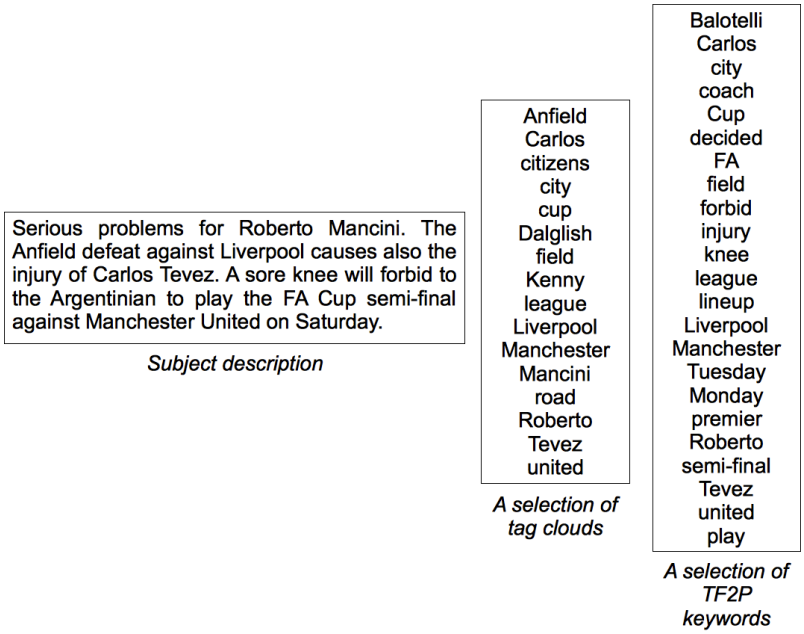

Fig. 1 An example of extracted keywords.

6 Conclusions and Future Work

In this paper, we presented a preliminary study aimed at verifying the effectiveness of adopting text summarization techniques to suggest keywords to news and news aggregations in a multimodal aggregation system. To perform our study, we compared ten different extraction-based techniques with the keywords provided by the adopted multimodal aggregation system. Results, calculated in terms of precision, recall, and F1, shown that the best performances are obtained when using the TF2P technique for both news and news aggregations. In other words, the best set of keywords is obtained considering the title, the first and second paragraph of each news.

As for the future work, we are setting up new experiments aimed at using further metrics to evaluate the effectiveness of the adopted techniques. In particular, we are studying how to adapt the approach to measure the keyword effectiveness index. Furthermore, we are setting up experiments aimed at investigating if merging the baseline keywords with those extracted by the most effective adopted text summarization technique lead to an improvement. Moreover, we are planning to select some users, asking them to give a degree of relevance to each keyword, e.g., relevant, somewhat relevant, or irrelevant.

References

1. Armano, G., Giuliani, A., Vargiu, E.: Experimenting text summarization techniques for contextual advertising. In: IIR'11: Proceedings of the 2nd Italian Information Retrieval (IIR) Workshop (2011)
2. Armano, G., Giuliani, A., Vargiu, E.: Studying the impact of text summarization on contextual advertising. In: 8th International Workshop on Text-based Information Retrieval (2011)
3. Baxendale, P.: Machine-made index for technical literature - an experiment. IBM Journal of Research and Development **2**, 354–361 (1958)
4. Bertini, M., Del Bimbo, A., Torniai, C.: Enhanced ontologies for video annotation and retrieval. In: Proc. of the 7th ACM SIGMM international workshop on Multimedia information retrieval, pp. 89–96 (2005)
5. Bertini, M., Del Bimbo, A., Torniai, C.: Automatic annotation and semantic retrieval of video sequences using multimedia ontologies. In: Proc. of the 14th annual ACM international conference on Multimedia, pp. 679–682 (2006)
6. Brandow, R., Mitze, K., Rau, L.F.: Automatic condensation of electronic publications by sentence selection. Information Processing Management **31**, 675–685 (1995)
7. Das, D., Martins, A.F.: A survey on automatic text summarization. Tech. Rep. Literature Survey for the Language and Statistics II course at CMU (2007)
8. Deschacht, K., Moens, M.F.: Finding the Best Picture: Cross-Media Retrieval of Content. In: Proc. of ECIR 2008, pp. 539–546 (2008)
9. Dimitrova, N.: Multimedia content analysis: The next wave. In: CIVR'03 Proceedings of the 2nd international conference on Image and video retrieval, pp. 9–18 (2003)
10. Edmundson, H.P.: New methods in automatic extracting. Journal of ACM **16**, 264–285 (1969)
11. Hanjalic, A.: Content-Based Analysis of Digital Video. Kluwer Academic Publishers (2004)
12. He, S., Tanaka, K.: Modeling omni-directional video. In: Advances in Multimedia Modeling, 13th International Multimedia Modeling Conference, MMM 2007, pp. 176–187 (2007)
13. Kokar, M.: Formalizing classes of information fusion systems. Information Fusion **5**(3), 189–202 (2004)
14. Kołcz, A., Alspector, J.: Asymmetric missing-data problems: Overcoming the lack of negative data in preference ranking. Information Retrieval **5**, 5–40 (2002)
15. Kolcz, A., Prabakarmurthi, V., Kalita, J.: Summarization as feature selection for text categorization. In: CIKM '01: Proceedings of the tenth international conference on Information and knowledge management, pp. 365–370. ACM, New York, NY, USA (2001)
16. Kraaj, W., Smeaton, A., Over, P.: TRECVID 2004: An overview. In: Proc. of TRECVID Workshop 2004 (2004)
17. Laudy, C., Ganascia, J.G.: Information fusion in a tv program recommendation system. In: 11th International Conference on Information Fusion, 2008, pp. 1–8 (2008)
18. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications **2**(1), 1–19 (2006)
19. Li, X., Yan, J., Deng, Z., Ji, L., Fan, W., Zhang, B., Chen, Z.: A novel clustering-based RSS aggregator. In: Proc. of WWW07, pp. 1309–1310 (2007)
20. Luhn, H.: The automatic creation of literature abstracts. IBM Journal of Research and Development **2**, 159–165 (1958)
21. Mahler, R.P.S.: Statistical Multisource-Multitarget Information Fusion. Artech House, Inc., Norwood, MA, USA (2007)
22. Messina, A., Bailer, W., Schallauer, P., et al., V.T.: Content analysis tools. Deliverable 15.4, PrestoSpace Consortium (2007)
23. Messina, A., Boch, L., Dimino, G., Allasia, W., et al., R.B.: Creating rich metadata in the tv broadcast archives environment: the prestospace project. In: Proc. of IEEE AXMEDIS06 Conference (2006)
24. Messina, A., Borgotallo, R., Dimino, G., Gnota, D.A., Boch, L.: Ants: A complete system for automatic news programme annotation based on multimodal analysis. In: Intl. Workshop on Image Analysis for Multimedia Interactive Services (2008)

25. Messina, A., Montagnuolo, M.: Information Retrieval and Mining in Distributed Environments, chap. Multimodal Aggregation and Recommendation Technologies Applied to Informative Content Distribution and Retrieval. A. Soro and E. Vargiu and G. Armano and G. Paddeu (2010)
26. Messina, A., Montagnuolo, M.: Heterogeneous data co-clustering by pseudo-semantic affinity functions. In: Proc. of the 2nd Italian Information Retrieval Workshop (IIR) (2011)
27. Nenkova, A.: Automatic text summarization of newswire: lessons learned from the document understanding conference. In: Proceedings of the 20th national conference on Artificial intelligence - Volume 3, pp. 1436–1441. AAAI Press (2005)
28. Nguyen, L.D., Woon, K.Y., Tan, A.H.: A self-organizing neural model for multimedia information fusion. In: 11th International Conference on Information Fusion, 2008, pp. 1–7 (2008)
29. Radev, D.R.: A common theory of information fusion from multiple text sources step one: cross-document structure. In: Proceedings of the 1st SIGdial workshop on Discourse and dialogue, pp. 74–83. Association for Computational Linguistics, Morristown, NJ, USA (2000)
30. Radev, D.R., Hovy, E., McKeown, K.: Introduction to the special issue on summarization. Computational Linguistic **28**, 399–408 (2002)
31. Snoek, C.G., Worring, M.: Multimodal video indexing: A review of the state-of-the-art. In: Multimedia Tools and Applications, pp. 5–35 (2005)
32. Wang, T., Yu, N., Li, Z., Li, M.: nReader: reading news quickly, deeply and vividly. In: In Proc. of CHI '06 extended abstracts on Human factors in computing systems, pp. 1385–1390 (2006)
33. Wu, Y., Chang, E.Y., Chang, K.C.C., Smith, J.R.: Optimal multimodal fusion for multimedia data analysis. In: MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 572–579. ACM, New York, NY, USA (2004)
34. Xu, C., Wang, J., Lu, H., Zhang, Y.: A novel framework for semantic annotation and personalized retrieval of sports video. IEEE Trans. on Multimedia **10**(3), 421–436 (2008)