

# 基于实体名的文本自动综述研究<sup>\*</sup>

葛加银 黄萱菁 吴立德

(复旦大学计算机科学与工程系 上海200433)

**摘要** 自动文摘是自然语言处理的一个重要分支,在信息检索领域中有着重要的用途,文本自动综述是自动文摘在多文档上的推广。本文提出了基于实体名扩展的自动综述方法,这种方法认为综述中的实体名个数反映其中所蕴含信息量的多少。我们用该方法实现针对事件的自动综述生成,并参加了2003年文本理解会议(Document Understanding Conference, DUC)进行统一评测, DUC 反馈的评测结果显示这种方法是有效的。此外,本文还对文本理解会议的任务、评测方法和测试结果做了简单介绍。

**关键词** 文本自动综述, 实体名, 自动文摘

## Automatic Text Summarization Based on Named Entity

GE Jia-Yin HUANG Xuan-Jiang WU Li-De

(Dept. of Computer Science and Engineering, Fudan University, Shanghai 200433)

**Abstract** Text Summarization is one important branch of Natural Language Processing and is very useful in Information Retrieval. This paper presents an approach of automatic text summarization based on Named Entity. This approach assumes that the number of Named Entities in a summary reflects the amount of information in it. By this approach we generate the summaries focused by events which are submitted to and evaluated by Document Understanding Conference 2003. The results of the evaluation show that this approach is effective. This paper also introduces the Document Understanding Conference.

**Keywords** Automatic text summarization, Named entity

## 0 引言

自动文摘是自然语言处理的一个重要分支,其目标为:从信息源中提取内容,采用压缩的形式和与用户(或应用)需求相关的方式,将最重要的内容呈现给用户<sup>[1]</sup>。按照文档数量,自动文摘可分为单文档自动摘要和多文档自动摘要,多文档自动摘要又称为文本自动综述。

随着 WWW 的出现,自动综述系统的研究具有重大的现实意义。作为信息检索研究的重要成果,不少搜索引擎已经投入实用,但它们还远不够完美。用户输入一个查询请求,搜索引擎返回大量相关文档。用户要较为全面地了解相关的重要信息,必须逐个浏览返回结果,不仅费时,而且还会因为大量的内容重复而感到厌倦。如果能将所有相关文档生成一篇综述,将大大方便用户。

文本自动摘要的方法可以粗略地分为两大类:浅层方法和深层方法。浅层方法利用一些统计特征和语言学特征进行文摘的自动生成,并不进行深层理解, J. Goldstein 等人在路透社和洛杉矶时报新闻语料上对各种特征做过充分的分析<sup>[2]</sup>。深层方法利用信息抽取和自然语言处理的技术,试图将文本转化为机器内部的语义表示形式,在“理解”了文本的基础上生成文摘,通常的语义表示形式是模板, D. Radev 等人1995年完成的自动文摘系统就使用了模板技术<sup>[3]</sup>。浅层方法是一种领域无关的方法,但浅层方法所生成的文摘质量相对较差。深层方法刚好相反,生成的文摘质量较好,但方法的具

体实现依赖于具体的领域,因为对于不同的领域,其内部的语义表示形式譬如模板往往是不一样的。

评测是自动文摘(包括自动综述)研究中的一大难点,由于缺乏可靠的第三方评测,自动文摘的算法很难得到客观的比较。近年来上述情况有所改变,国际上召开的一些评测会议为自动文摘提供了统一的评测平台,其中最著名的是文本理解会议<sup>[4]</sup>(DUC, Document Understanding Conference)。DUC 是美国国防部 DARPA TIDES (Translingual Information Detection, Extraction and Summarization program)项目的一部分,它由 NIST (美国国家标准技术局)具体管理执行。第一届会议开始于2001年11月,至今已进行了三届会议。参加单位包括:BBN、NTT、密歇根大学、哥伦比亚大学、纽约大学、剑桥大学等。这个会议除了像别的学术会议一样交流论文外,还组织对各个单位的系统进行评测。

复旦大学作为第一批国内的研究机构参加了2003年第三届国际文本理解会议,完成了这次会议提出的主要评测任务:针对事件的短文摘生成。具体而言,给系统三十个 TDT (Topic Detection and Tracking, 另一个关于主题检测和跟踪的国际评测会议<sup>[5]</sup>)文档集和三十个与文档集一一对应的主题,要求系统为每个文档集生成100单词左右的综述,这个综述必须与对应话题相关。

我们提出了一种新方法,该方法采用抽句的方式进行自动综述生成,所以对句子的分析是工作的重心所在。我们认为,一个句子被抽取成为针对事件的综述,必须具备以下三个

<sup>\*</sup> 本文受国家自然科学基金项目(69935010, 60103014)、863项目(2001AA114120, 2002AA142090)资助。葛加银 硕士研究生,主要研究方向为自动文本摘要。黄萱菁 博士,副教授,主要研究方向为自然语言处理与信息检索。吴立德 教授,博士生导师,主要研究方向为图像处理、模式识别、计算机视觉和自然语言处理。

条件:(1)概括性强,信息量大。这是成为文摘句的必要条件;(2)与对应的话题所描述的事件相关;(3)与其他被抽取的句子之间信息冗余较小。这一点对于自动综述非常重要,因为关于某个主题的一组文章在内容上往往有很大的冗余度。

任何事件都会牵涉到一些人物、组织和地区,与事件相关的信息大多可抽象为这些人物、组织和地区的行为、状态或它们之间的关系,因此这些与事件相关的实体名(Named Entity)有助于判断一个句子是否满足以上条件。含有较多相关实体名的句子一般也含有较大的信息量,且与事件相关;如果两个句子含有不同的实体名,它们之间的信息冗余一般较小。我们从话题以及相关的文档中提取出与事件相关的实体名,并利用这些实体名抽取合适的句子作为综述。这种方法抓住了构成信息的要素之一:实体名,并利用实体名判断句子所含信息量的多少,与事件的相关程度,以及句子间的信息冗余程度。相对于一般的浅层方法而言,这种方法对文本信息作了较为深入的处理。同时,这种方法的实现独立于具体的领域,这又有别于一般的深层方法。所以这种方法介于浅层和深层两类方法之间,并结合了这两类方法的优点。DUC 返回的评测数据显示这种方法是行之有效的。

本文第2节介绍 TDT 话题的预处理,主要是实体名提取;第3节具体介绍两种算法,并对它们进行比较;第4节介绍 DUC 的评测方法和我们提交的结果在 DUC 评测中取得的成绩;最后是总结和展望。

## 1 TDT 话题预处理

在具体说明预处理工作之前,先介绍一下 TDT 话题。一个 TDT 话题描述一个或若干个相关事件,以及其背景、影响和相关评论。每个话题有一个标题和一个种子事件(Seminal Event),种子事件包括 WHAT(什么事)、WHO(主要人物或者组织机构)、WHEN(发生时间)和 WHERE(发生地点);每个话题还有话题说明(Topic Explication),以自然语言的方式对该话题做较为详细的描述。对 TDT 话题的预处理主要针对标题和种子事件。

一个话题一般有一个种子事件,而话题的种子事件会涉及到一些实体名,包括:人物、组织、国家或地区。在相关的事件以及背景、影响和评论中,这些实体名也会出现。因此,一个能较好地概括相关信息的综述应该涉及这些实体名。从上面的介绍中可知,TDT 话题的描述中也包括了这种与话题密切相关的实体名,我们将这些实体名提取出来。

信息抽取的研究已提出了不少实体名提取的方法,在这里实体名提取是通过专有名词短语识别来实现的。首先对文档中每个词做词性标注,进而找到所有名词短语,然后根据大小写、词频等特征识别出专有名词短语。同一个实体名可能会以不同的专有名词短语形式出现,所以我们对这些专有名词短语进行聚类,每一个类对应一个实体名。

从话题中提取出来的实体名与话题密切相关,所以有助于判断一个句子是否相关,即是否满足条件(2),同时,它们还能用来判断一个句子是否满足条件(1),一般而言,如果一个句子包含多个实体名,即该句所述内容涉及多个实体名,它的信息量应该是较大的。

与事件相关的信息会涉及相关的实体名,但实体名本身不能构成完整的事件信息。譬如:“杨利伟”和“香港学生”本身并不包含事件信息。而句子“杨利伟与香港学生真情对话”指出了这两个实体名之间发生的关系:对话,这才给出完整的事件信息。除了这种情况外,实体名与其行为、状态也可构成完整的事件信息,如“神舟五号载人飞船成功发射升空”。

虽然我们的系统没有对文章进行深层理解,实际情况也往往比上面的例子复杂,但在对信息构成进行上述分析的基础上可以做一些相关的浅层处理。我们将话题的标题和种子事件描述中的实义词抽取出来作为关键词,这些关键词在话题中往往用来说明实体名的行为、状态或者实体名之间的关系,这些信息与话题密切相关,所以这些关键词可以用来判断一个句子与话题的相关性,即满足条件(2)的程度。

从话题中抽取出来的实体名数量一般比较少,因此我们把文档集内相关句子中的实体名(不包括同时在话题中也出现了的实体名)也抽出来作为补充。为区别起见,从话题中抽取出来的实体名称为话题实体名(TNE, Topic Named Entity),而从文档集中抽取出来的实体名称为文档实体名(DNE, Document Named Entity)。为了得到尽量多的 DNE,只要一个句子含有至少一个关键词,我们就认为其相关,进而从中抽取实体名。这样做会不可避免地引入噪声,我们对 DNE 在文档集中的出现频数进行了统计,根据统计的结果我们可以较好地解决噪声问题。

前面已经说过,实体名可以用来判断一个句子是否满足条件(1)和(2),其实,在判断一个句子是否满足条件(3)的时候,实体名也非常有用。如果两个句子含有不同的实体名,那么它们之间信息冗余往往较少。

## 2 算法实现、对比和分析

我们设计了两种不同的算法来利用实体名实现针对事件的自动综述。本节详细介绍这两种算法,并对它们进行对比和分析。为叙述方便,首先给出几个定义。

对每个文档集及其对应的话题,我们定义三个集合:  $KWSet$ 、 $TNESet$  和  $DNESet$ ,分别由 TDT 话题预处理后得到的关键词、TNE 和 DNE 组成。我们还将文档集中所有未被抽取到综述中的句子组成集合  $SenSet$ ,已被抽取至综述中的句子组成集合  $ExtSet$ 。开始时, $SenSet$  包含文档集中所有的句子, $ExtSet$  为空,每抽取一个句子至综述中,这两个集合会发生变化。对文档集中的每个句子  $s$ ,我们从中抽取出实义词组成集合  $WSet(s)$ ,从中抽取出实体名组成集合  $NESet(s)$ 。

### 2.1 基于实体名的简单加权算法

对  $SenSet$  内的每个句子  $s$  根据下式得到评分  $Score(s)$ :

$$Score(s) = |KWSet \cap WSet(s)| + \alpha \cdot |TNESet \cap NESet(s)| + \beta \cdot \frac{\sum_{NE \in NESet(s) \cap DNESet} NE.freq}{\max_{NE \in DNESet} NE.freq} \quad (1)$$

其中  $NE.freq$  为实体名 NE 在文档集中出现的频数。这里有两个参数,在我们的系统中, $\alpha$  取值为 2.5, $\beta$  取值为 1。

句子含关键词的多少往往反映其与话题的相关程度;句子含话题实体名的多少既反映该句与话题的相关程度,又反映其所含信息的多少;文档实体名在考虑出现频数的条件下也能较好地起到话题实体名的作用。式(1)中的三项分别反映了句子在这三方面的情况,所以,那些既与话题相关又有较大信息量的句子,即同时满足条件(1)和(2)的句子可以根据式(1)得到较高的评分。因此,得分最高的句子  $s_1$  被抽取出来

$$ExtSet \leftarrow ExtSet \cup \{s_1\}$$

$$SenSet \leftarrow SenSet \setminus \{s_1\}$$

下一个被抽取的句子除了仍然需要满足条件(1)和(2)之外,还必须给综述带来  $s_1$  所没有的新信息,以保证综述的信息量尽量大。一般而言,这种句子应该含有  $s_1$  所未包含的实体名,所以我们从  $TNESet$  和  $DNESet$  中去掉  $NESet(s_1)$  中所含实体名,即

$$TNESet \leftarrow TNESet \setminus NSEt(s_1)$$

$$DNESet \leftarrow DNESet \setminus NSEt(s_1)$$

然后对  $SenSet$  中的所有句子根据式(1)重新打分,得分最高的句子  $s_2$  较好地满足三个条件,被抽取进综述。对集合做相应的更新,继续抽取新的句子,直至综述达到指定大小。

## 2.2 实体名稳步扩展算法

在前面的算法中,一个重要的原则是每抽取一个新的句子,就必须加进新的实体名,从而加进新的信息。在抽取了三句之后, $TNESet$  中将会剩下很少的话题实体名,而  $DNESet$  中所剩文档实体名的重要性也会大打折扣。尽管按照这种算法生成的综述最终含有较多的实体名,但噪声也可能随之引入。针对这一点,我们提出一种新的算法,算法仍然扩展综述所包含的实体名,但不是一味盲目地扩展。在第一句被抽出之后,算法要求每次新抽取的句子必须包含新的实体名,对所有满足这一条件的句子进行评分,并从中抽取得分最高的句子,这里的评分不再考虑句子含新实体名的多少,仅考虑句子含关键词的多少。按照这种方法,最终生成的综述仍然会含有较多的实体名,而且这些实体名的重要性也得到了保证。这个算法每抽取新的句子都是较为稳当地而不是盲目地增加综述中的实体名,所以这个算法被称为实体名稳步扩展算法。下面予以具体介绍。

前面已经讨论过,依据式(1),较好满足条件(1)和(2)的句子可以得到较高的评分。因此,我们从文档集中选择得分最高的若干句子组成集合  $CandiSenSet$ ,综述中的句子将从这个集合中抽取出来,这个集合的大小取决于综述的指定大小。

仍然选择根据式(1)得分最高的句子  $s_1$  作为第一个被抽取的句子。与此同时,从  $TNESet$  和  $DNESet$  中去掉  $s_1$  中所含实体名。如果此时  $TNESet$  中剩下的实体名较少,从  $DNESet$  中取出在文档中出现频数最高的实体名补充进  $TNESet$ ,即

$$\begin{cases} TNESet \leftarrow TNESet \cup \{ \arg \max_{NE \in DNESet} NE.freq \} \\ DNESet \leftarrow DNESet \setminus \{ \arg \max_{NE \in DNESet} NE.freq \} \end{cases}$$

$TNESet$  必须含有  $StdSetSize$  个实体名,这里  $StdSetSize$  是参数,在我们的系统中取为6。

对  $CandiSenSet$  也作相应的修改:

$$CandiSenSet \leftarrow CandiSenSet \setminus \{s_1\}.$$

在对  $CandiSenSet$  中的句子进行评分之前,先确定下一个被抽取的句子至少须含有多少话题实体名,这个下限值我们用  $MinTNECount$  表示,具体确定办法如下:

$$MinTNECount = \min(2, \max_{s \in CandiSenSet} |NSEt(s) \cap TNESet|) \quad (2)$$

在此基础上我们对  $CandiSenSet$  中的每个句子  $s$  按照下式进行评分

$$NewScore(s) = \begin{cases} 0 & |NSEt(s) \cap TNESet| < MinTNECount \\ |WSet(s) \cap KWSet| & \text{otherwise} \end{cases} \quad (3)$$

我们选择得分最高的句子  $s_2$  作为第二个被抽取的句子,并作相应的集合更新,如此进行下去,直至由这些被抽取的句子所组成的综述达到指定大小。

式(3)仅仅将句子是否有足够的新话题实体名作为必要条件,对于所有满足该条件的句子而言,其得分高低与其所含新话题实体名的多少无关,仅仅取决于其所含关键词的数目,这种评分方式可能偏于简单,所以我们最开始用式(1)对文档中的每个句子进行打分,选择高分句子组成集合  $CandiSenSet$ ,后面的所有工作都是在这个集合上进行,以防止将不满足条件(1)和(2)的句子抽取到综述中来。另外,式(3)仅仅考虑  $TNESet$ ,只有在  $TNESet$  中实体名较少时才从  $DNESet$  中选择出现频数最高的若干实体名作为  $TNESet$  的补充。这样做充分利用了话题实体名,合理利用了文档实体名,有效避免了噪声。

## 2.3 对上述算法的对比和分析

我们对上述两种算法进行了人工比较。具体方法如下,对每个话题,评测人员先阅读 TDT 的话题说明,然后阅读两种算法所生成的综述,综合考虑两方面对其进行评分:1)信息量的大小;2)是否与话题相关,有没有与话题无关的信息。共有两名评测人员参加了评测工作,两人一致认为:实体名稳步扩展算法的性能要优于基于实体名的简单加权算法。这个结论说明我们前面的分析是正确的,稳步增加文摘中的实体名可以避免引入噪声。

我们还在 DUC2002 提供的文摘语料上研究句子被抽取成为综述的条件,对文档集中的句子提取有关特征,在此基础上训练前馈神经网络(BPNN)<sup>[6]</sup>。我们利用训练好的 BPNN 生成针对事件的综述,并将其与以式(2)和(3)所得结果进行比较,我们发现实体名稳步扩展算法在性能上稍好于 BPNN。这说明,基于实体名的文本自动综述方法是有效的,这一点在 DUC 评测中得到进一步的证实。

## 3 评测

### 3.1 DUC 评测方法

自动文摘的评测方法可以分为两大类:内部评测和外部评测<sup>[6]</sup>。前者是就一个独立的文摘系统以某些性能标准对其本身进行评测;后者则是在一组系统中,在文摘系统和其它系统如检索系统、问题回答系统等进行相互作用的情形下,通过考察文摘系统与外部环境之间的联系进行评测。DUC 目前采用的是内部评测方法。

为了对各单位提交的针对事件的综述进行评测,DUC 为每个文档集提供了4份人工综述和两份简单的自动综述作为 baseline,其中人工综述由专业评测人员手工生成,自动综述由两种不同的简单方法自动生成。在4份人工综述中,有一份作为标准综述(model)被用来对所有单位提交的综述、简单的自动综述以及其他3份人工综述(称为 peers)进行评测。

评测工作分两方面:综述的表达质量;综述的内容。对表达质量进行的评测独立于标准综述,主要是看综述本身有没有语法错误,是否通顺,可读性如何。评测人员在通读综述后,回答12个问题,每一个问题都针对一类表达错误,如:

About how many gross capitalization errors are there?

这个问题针对大小写错误。对每个问题,评测人员根据问题所针对的表达错误在综述中出现的情况对综述进行评分,共有4个级别:{0,1-5,6-10,more than 10}。如果某个综述中没有出现问题所针对的错误,评测人员给出的评分就为0。通过这种方法,得到两个指标:评分为0的问题数目(Count of Quality questions with non-0 answers,记为 CQ)和各问题的平均评分(mean of the Quality question Scores,记为 QS)。

对综述内容的评测是通过与标准综述的对比进行的。首先将 Model 分割成若干反映基本信息的单元 MU (Model Unit),对综述也作这样的分割得到若干 PU (Peer Unit)。标记出这样的 PU:其反映的信息被某个或者某几个 MU 反映。显然,所有被标记的 PU 放在一起所反映的信息皆为 Model

所反映,但只占 Model 所反映信息的一部分,下面两个指标反映了这一部分所占比例的大小:信息覆盖比例平均值(Mean Coverage,记为 MeanC)和信息覆盖比例中位值(Median coverage,记为 MedianC)。考虑综述的长度又得到两个指标:考虑长度的信息覆盖比例平均值(Mean Length-adjusted Coverage,记为 MeanLC)和考虑长度的信息覆盖比例中位值(Median Length-adjusted Coverage,记为 MedianLC)。未被标记的 PU 所反映信息未直接出现于 Model 中,但可能会有部分信息与标准综述所反映的主题相关,因此又得到一个指标:至少与标准综述主题相关的 PU 在所有未被标记的 PU 中所占比例(Fraction of Unmarked peer units at least Related to the model's subject,记为 FUR)。

### 3.2 评测结果

2003年共有16家单位提交了针对事件的综述,DUC 评测人员根据上面介绍的方法对所有的综述进行了统一评测,表1给出了评测结果。在表1中,SysNum 是 DUC 给各个自动综述系统的编号,我们系统的编号为12,DUC 的两个简单自动综述系统分别编号为2和3。没有作为标准综述的人工综述也一

起参加了评测,Human 一栏为其各项指标评测结果的均值。我们还分别按照各个指标对所有参评单位的结果进行了排名,对于每个指标,表1在给出了我们系统的评测结果之后,在括弧中给出根据这一指标我们在16家参评单位(不包括人工评测和 DUC 的两个简单系统)中的名次。

从表1中可以发现,我们的系统在表达质量方面性能较好,排名较为靠前。在内容方面大部分指标居于中上游,因为我们的综述较长,所以在考虑长度的时候我们的排名有所下降。我们还注意到,根据指标 FUR 我们的系统排名为4,一般而言,这一项指标对于较长的文摘是不利的。因为绝大多数系统在生成综述的时候,尽量给出最重要最相关的信息,综述越长,不相关信息所占比例一般会越高,所以在未被标记的 PU 所反映的信息中,与标准综述主题相关的信息所占比例一般会较低。

我们所提交的综述是由2.2节介绍的算法生成的,这个算法其实并不复杂,但性能在所有参评单位中居于中上游,这说明基于实体名进行自动综述生成的方法是有效的,可以按照这个思路把工作深入下去。

表1 DUC 评测结果

SysNum	CQ	QS	MeanC	MedianC	MeanLC	MedianLC	FUR
Human	0.62605	0.361345	0.470328	0.47605	0.303261	0.306655	0.918824
2	1.32203	0.762712	0.157305	0.00678	0.090746	0.003729	0.742857
3	2.84746	1.0339	0.187915	0.030509	0.115407	0.019932	0.775
6	1.40678	0.813559	0.274644	0.132203	0.180729	0.087305	0.8
10	1.80682	0.897727	0.182523	0.023864	0.143295	0.028909	0.888889
11	0.966102	0.525424	0.196661	0.061017	0.147864	0.045627	0.692308
12	<b>1.57827 (5)</b>	<b>0.79881 (3)</b>	<b>0.24178 (7)</b>	<b>0.111864 (5)</b>	<b>0.143188 (11)</b>	<b>0.064138 (7)</b>	<b>0.878923 (4)</b>
13	3.15254	1.02824	0.288949	0.118644	0.188254	0.076949	0.706667
14	2.54237	0.917525	0.245746	0.084746	0.173763	0.062559	0.7625
15	7.86441	1.33631	0.075254	0.00678	0.054102	0.005322	0.420339
16	1.27119	0.830508	0.272051	0.125424	0.176492	0.085864	0.916667
17	1.81356	0.915254	0.164525	0.027119	0.095373	0.015322	0.68
18	2.57627	0.932203	0.220797	0.074576	0.151322	0.049898	0.662222
19	2.44068	0.966102	0.142322	0.054237	0.100525	0.034441	0.6
20	2.32203	1.08475	0.259898	0.172881	0.163864	0.109915	0.717647
21	2.64407	0.97739	0.181373	0.040678	0.12278	0.026576	0.568421
22	1.20339	0.627119	0.215169	0.071186	0.17678	0.064136	1
23	2.28814	0.959051	0.248593	0.071186	0.174254	0.052712	0.8
26	1.94915	0.949153	0.218814	0.074576	0.141322	0.049525	0.858333

**总结与展望** 本文首先提出句子被抽取成为针对事件的综述必须满足的三个条件,与事件相关的实体名有助于判断一个句子是否满足以上条件。我们分别从 TDT 话题和与话题相关的句子里提取出话题实体名和文档实体名,另外,在 TDT 话题中描述实体名行为、状态或者实体名之间关系的关键词也被提取出来。在此基础上设计了两个不同的算法,每个算法在抽取新的句子时都力图增加新的实体名。对这两个算法进行的比较显示,稳步地增加文摘中的实体名,可以避免引入噪声。我们提交了实体名稳步扩展算法生成的综述参加 DUC 统一评测,评测结果显示,该算法的性能在所有参评系统中居于中上游。实体名稳步扩展算法是比较简单的,但基于实体名进行自动综述生成的方法是有效的,具有一定的探索价值,我们将来要做的工作就是按照这个思路将我们的工作深入下去。我们可以采用更有效的实体名提取方法,并在此基础上实现一种简单但通用的机器内部表示语义的形式,可以帮助系统决定抽取哪些句子使得生成的综述信息量最大且与

话题相关。

### 参考文献

- 1 Mani I. Automatic Summarization. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2001
- 2 Goldstein J, Kantrowitz M, Mittal V O, Carbonell J. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. In: Proc. of SIGIR-99, Berkeley, CA, Aug. 1999
- 3 McKeown, Kathleen R, Radev D R. Generating Summaries of Multiple News Articles. In: Proc. of ACM Conference on Research and Development in Information Retrieval, SIGIR'95. Seattle, WA, July 1995
- 4 <http://tides.nist.gov/>
- 5 <http://www ldc.upenn.edu/Projects/TDT3/topics.html>
- 6 Baldwin B, et al. An evaluation road map for summarization research. TIDES, July 2000