문서 자동 요약의 현황과 과제*

강인수**

<차 례>

Ⅰ. 서론

컴퓨터와 인터넷의 영향으로 전자화되어 기록, 유통되는 정보의 양은 기하급수적으로 증가하고 있으나 원하는 정보를 찾기 위한 인간의 요구를 충족시키는 일은 더 어려워지고 있다. 한 연구[1]에 따르면 10년 전 한 해(2002년) 인쇄/필름/자화/광학 매체에 기록되는 새로운 정보의 양은 5엑사바이트(Exabyte=10^{18})이며 이 중 92%가 하드디스크와 같은 자기 매체에 저장된다고 한다. 5엑사바이트는, 미국 의회도서관에 소장된 1천 7백만 여 권의 도서들이 136테라바이트(terabyte=10^{12})로 전자화된다고 할 때, 미 의회 도서관 3만 7천여 개에 해당하는 정보의 양이다. 또한 2010년 8월 현재,

* 이 논문은 국어교육학회 제46회 발표대회(2010. 8. 28)에서 발표한 내용을 수정보완한 것이다(이 논문은 2010학년도 경성대학교 학술연구비지원에 의하여 연구되었음).
** 제1저자(교신저자), 경성대학교 컴퓨터학부 교수
1) http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/

웹(World Wide Web)은 최소 270억 웹 페이지들을 담고 있으며[2] 이로부터 구글 검색엔진은 질의어 'summarization'에 대해 5백 8십만 웹 페이지를 검색 결과로 제시하고 있다.

이처럼 끊임없이 그 양이 증가하는 초대용량 정보 저장소로부터 원하는 정보를 찾고자 하는 인간의 정보 요구를 만족시키기 위해 검색, 추출, 요약, 질의/응답 등 다양한 작업들을 자동화하는 도구들이 개발되어 사용되고 있으나 그 성능은 아직 인간을 만족시키기에 많이 부족하다. 이 중 요약은 문서(들)를 입력으로 받아 그 내용의 주요 핵심을 간략하게 정리하는 것으로 검색, 추출보다 고수준의 인간 친화적 작업에 해당한다. 본 논문은 문서 요약을 자동화하는 기술의 현황과 향후 과제들을 다룬다.

Ⅱ. 전문용어[3]

표준국어대사전에 따르면 '요약'은 "말이나 글의 요점을 잡아서 간추림"이며 '간추림'은 "글 따위에서 중요한 점만을 골라 간략하게 정리하다"로 정의된다. 웹 상의 협업 백과사전인 위키피디아에서는 'automatic summarization'을 "creation of a shortened version of a text by a computer program"으로 정의하고 있다. 자동 요약 연구자들에게 'text summarization'은 "the process of distilling the most important information from a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)"(Mani & Maybury, 1999)로 받아들여진다.

자동 문서 요약의 사용 환경은 입력의 형태, 출력의 형태, 의도/용도(purpose)에 따른 세 가지 인자(factor)로 구분하는 것이 일반적이다. 요약 시스템의 입력 텍스트가 하나의 문서인 경우를 단일 문서 요약(Single-

2) http://www.worldwidewebsize.com/

3) 이 절의 많은 부분들은 (Mani & Maybury, 1999 ; Sparck-Jones, 1999)의 내용을 인용한 것이다.

Document Summarization, SDS)으로, 특정 주제와 관련된 여러 문서인 경우를 다중 문서 요약(Multi-Document Summarization, MDS)으로 정의한다. 다중 문서 요약의 한 형태인 질의 관련 다중 문서 요약(Query-Focused Multi-Document Summarization, QFMDS)은 현시대 정보 검색/분석 작업의 전형적 유형 중 하나이다. 이는 사용자가 작성한 질의문(query, topic, information need)에 대해 검색 엔진을 통해 검색된 상위 수십 건의 관련된 다중 문서의 내용을 요약하여 사용자에게 제시하는 것이다.

요약 시스템의 출력 텍스트와 관련하여 그 형태가 입력 텍스트의 발췌문(extract)들에 의존하는 경우를 추출 요약(extractive summarization, extraction)으로, 새롭게 작문된 문장들로 구성된 경우를 추상 요약(abstractive summarization, abstraction)으로 정의한다.

독자층에 대한 요약의 의도와 관련하여, 광범위한 독자를 대상으로 작성된 요약문을 포괄적 요약문(generic summary)으로, 특정 독자(의 관심 영역)를 염두에 두고 작성된 요약문을 사용자 지향적 요약문(user-focused summary)으로 정의한다. 포괄적 요약문은 불특정 다수의 독자를 대상으로 저자의 관점에서 원 저자나 요약 전문가가 작성하는 것이며, 사용자 지향적 요약문은 특정 사용자의 정보 요구(예 : 질의문)와 관련된 주제의 관점에서 요약문을 작성하는 것에 해당한다. 후자의 경우 하나의 문서는 복수 개 주제(topic)를 다루며 각 주제에 대해 여러 측면(aspect)을 기술하고 있다는 가정이 전제된 것이다. 요약문이 입력 텍스트의 내용을 대신하는 용도를 가진 경우 정보적 요약문(informative summary)이라 하며, 입력 텍스트가 어떠한 주제를 다루는 것인지를 알리기 위한 용도를 가진 경우를 지시적 요약문(indicative summary)이라 정의한다. 예를 들어, 학술 문헌의 제목과 초록은 각각 지시적, 정보적 요약문에 해당할 수 있다.

요약율(compression ratio)은 요약 대상이 되는 입력 텍스트에 대한 출력 요약문 길이의 비율로 정의되며, 자동 문서 요약에서는 주로 30% 이내의 요약율을 갖는 요약문 생성을 다루어 왔다. 요약율이 감소할수록 요약문에 담긴 정보의 양은 줄어든다. 한 연구에서는 요약문이 정보적(informative)이기 위해서는 최소 20% 이상의 요약율을 가져야 한다고 보고하였다

(Morris et al., 1992).

Sparck-Jones(1999)와 Mani 등(1999)은 자동 문서 요약(automatic text summarization)을 입력 텍스트의 해석(interpretation), 해석된 내용을 내부 요약 표현으로의 변환(transformation), 내부 요약 표현으로부터 요약문[4] 생성(generation)의 세 단계로 구분하였다. Hovy(2005)는 이를 주제 확인(topic identification), 해석(interpretation), 요약문 생성(summary generation)의 세 단계로 나누었다. 전자는 보다 이상적인 추상 요약의 관점이 강하며 후자는 보다 현실적인 추출 요약의 관점이 강하다. 현재까지 연구되고 개발된 대부분의 요약 기법들이 추출 방식에 의존하고 있음을 고려하여 이 논문에서는 전체적으로 후자의 구분을 따른다. 그러나 추상 요약에 대해서는 전자의 단계들을 사용하여 기술한다.

Ⅲ. 문서 자동 요약

1. 추출 요약

1) 개요

추출 요약은 입력 문서에 대한 주제 확인(topic identification)을 거쳐 주요 문장들을 발췌하고, 발췌문의 길이를 줄이거나 발췌문 나열의 부자연스러움을 제거하는 요약문 생성(summary generation) 과정을 거치는 것이 일반적이다. 즉 자동 요약의 세 단계 중 두 번째 단계인 해석이 생략된다. 보다 단순한 추출 요약 시스템의 경우 세 번째 단계가 생략되기도 한다. 주제 확인을 거쳐 추출된 문장들의 나열은 대용어의 참조어 누락, 명제의 중복, 담화 표지의 부적절성 등의 부자연스러움이 빈번히 발견되므로 요약문 생

4) 요약문의 단위는 문장이 일반적이나 구(phrase)나 키워드(들)일 수도 있다.

성 단계에서 이를 해소하는 절차를 수행하게 된다. 요약문 생성 단계에서 문장 압축(sentence compression)이 적용되기도 하는데 이는 한 문장 내에서 시간 표현, 내포절, 동격구 등의 부가 표현들을 제거하는 과정이다(Zajic et al., 2007).

추출 요약에서 사용된 주제 확인 기법들은 주제 빈출성, 주제 지역성, 담화 구조 주제 유추성, 인간 요약 모방성, 요약문 최소 중복성 가정 등에 기반하여 입력 텍스트의 주제 어구 및 문장을 찾는다. 주제 빈출성(topic dominance, topic prevalence)은 문서의 주제와 관련된 용어들은 빈번히 출현한다는 가정(Luhn, 1958)으로, 문서 내에서 자주 등장하는 내용어들이 문서의 주제를 대표할 수 있다는 접근법으로 연결되며, 가장 전통적이며 대표적인 주제 확인 원칙 중 하나이다. 주제 지역성(topic locality)은 문서의 주제문들은 문서 내의 특정 위치(예 : 문서의 첫 단락 혹은 마지막 단락)에 나타난다는 가정(Baxendale, 1958 ; Edmundson, 1969)으로 문서의 내용에 무관하게 문서의 구조로부터 주제 문장의 위치를 결정하려고 하는 접근법에서 사용된다. 담화 구조 주제 유추성은 글의 담화 구조(discourse structure)로부터 핵심 문장들을 파악할 수 있다는 가정이다. 인간 요약 모방성은 사람이 작성한 요약문들로부터 주제문 선별의 자질을 모방할 수 있다는 가정으로, 기계학습을 사용하는 요약 시스템들은 모두 이 가정을 사용하는 것이다. 요약문 최소 중복성은 요약문을 구성하는 주제 문장들은 상호 간 내용의 중복이 최소화되어야 한다는 가정으로 특별히 다중 문서 요약에서 많이 사용된다.

2) 주제 빈출성

주제 빈출성에 기반하여 주제 (문장) 확인을 시도하는 접근법들은 주제 빈출성을 판단하기 위한 문서 표현으로 용어 집합, 어휘 그래프, 문장 그래프 등을 사용하였다. 용어 집합 방법은 문서를 출현 용어들의 집합으로 가정하고 문장 내 출현 용어의 주제성(topicality)으로부터 문장의 주제성을 계산하여 주제성이 큰 문장을 요약문으로 추출하는 시도이다. 자동 문

서 요약의 최초의 시도로 인용되는 Luhn(1958)의 연구에서는 학술 문헌의 요약을 위해 중간 빈도수를 갖는 내용어(content word)들을 주제어(significant words)로 가정하고 많은 주제어들이 한 문장 내에서 근접하여 출현하는 정도를 문장의 주제성으로 계산하였다.

Luhn과 이후의 많은 연구자들은 주제 빈출성을 실현하기 위해 문헌 내 용어 빈도(Term Frequency, TF)와 함께 역 문헌 빈도(Inverse Document Frequency, IDF)를 동시에 고려하는 주제성 계산법을 시도하였으나 근본적으로 어형(word form)의 출현 빈도로부터 용어의 주제성을 계산하고자 하였다. 이 방법이 간단하여 구현이 쉽고 계산 복잡도가 크지 않은 장점이 있으나 심층 수준의 주제 빈출성을 충실히 반영하지 못하는 단점이 있다. 이는 자연어 텍스트 내에서 주제 관련 어휘들은 동일 어형의 단순 반복 이상으로 동의어, 말바꿈(paraphrase) 표현, 연어(collocation), 대용어(anaphora), 상/하위어, 전체/부분어 등으로 출현하거나 생략되기도 하기 때문이다. 이 뿐 아니라 어형에 기초한 집계 방식은 동형이의어(homograph), 다의어(polyseme)가 출현한 경우 주제어의 빈도수를 각각 과다, 과소 집계하는 오류가 발생하기도 한다.

따라서 주제 빈출성을 최대한 활용하기 위해서는 어휘가 아닌 의미나 개념 단위의 주제 표현이 필요하며 생략어나 대용어의 올바른 참조어 또한 미리 결정해 두어야 한다. 전자의 경우 어의중의성해소(Word Sense Disambiguation, WSD) 기법이 요구되며 후자의 경우 대용어 참조 해결 기법(anaphora resolution)이 필요하다. 영어의 경우 동의어, 상/하위어, 전체/부분어의 처리를 위해서는 워드넷(WordNet) 등의 어휘개념망을 활용할 수 있으며 연어, 공기어(co-occurring term) 등의 관련어들은 대용량 말뭉치 분석을 통해 미리 획득된 어휘지식을 이용할 수 있다.

어휘 그래프 방법은 입력 문서에 대해 그 출현 용어를 노드(node, vertex)로 용어 간 관계를 노드 간 링크(link, edge)로 설정한 용어 그래프(망, network, graph)로 표현해 두고 용어 그래프의 구조로부터 문장의 주제성을 결정하는 시도이다. Barzilay 등(1997)은 어휘 그래프의 한 형태인 어휘 사슬(lexical chain)을 문서 요약에 적용하였다. 어휘 사슬은 텍스트가 갖는 응

집성(cohesion) 유형 중 식별이 용이한 반복, 동의어, 상위어, 공기어 등의 어휘적 응집성(lexical cohesion) 관계로 연결된 어휘들의 집합을 의미한다(Halliday & Hasan, 1976). 어휘 사슬에 기반한 주제 확인 방법은, 입력 텍스트에 대해 모든 가능한 어휘 사슬들을 생성하고 어휘들 간 연결 조밀성이 높은 상위 n개의 각 어휘 사슬과 관련된, 문서 내 첫 문장을 주제문으로 추출하는 것이다(Barzilay & Elhadad, 1997). 이 경우 주제 빈출성은 가장 조밀한 연결을 갖는 어휘 사슬들로 표현된다. Barzilay 등(1997)은 품사 태깅, 부분 파싱(partial parsing, shallow parsing)을 거쳐 분석된 문서 내 명사 상당어구에 대해 워드넷을 활용한 의미 관계들로 어휘 사슬을 만들어 요약에 적용하였다.

문장 그래프 방법은, 문서 내 문장들을 노드로 문장 간 유사도를 노드 간 링크로 설정하여 만들어진 문장 그래프에 대해 문장(노드) 순위화 알고리즘을 수행한 다음 상위 n개 노드에 해당하는 문장들을 요약문으로 추출하는 방식이다. 문장 순위화를 위한 가정은 문장 그래프 내에서 다른 많은 문장(노드)들과 강한 연결(링크)을 맺고 있는 문장(노드)일수록 문서 내에서 더 중요하다는 것이다. 이 가정을 단순히 실현한 것이 노드의 차수(degree[5])가 높은 순으로 요약문을 추출하는 방식으로 차수 중심성(degree-centrality) 기법이라 불린다(Salton et al., 1997 ; Erkan & Radev, 2004). 그러나 차수 중심성 기법은 한 문장과 관련(연결)된 다른 문장들의 중요도를, 실제 그렇지 않음에도 불구하고, 모두 동일하게 고려하는 단점이 있다.

이를 극복하는 기법이 문장 그래프에 대해 PageRank(Brin & Page, 1998), HITS(Kleinberg, 1999)와 같은 노드 순위화 알고리즘을 수행한 다음 상위 n개 노드에 해당하는 문장들을 요약문으로 추출하는 방식이다. PageRank는 하이퍼링크로 상호 연결된 웹 페이지들의 중요도(rank)를 계산하는 알고리즘으로 구글 검색 엔진에 성공적으로 적용되었다. 문장 그래프에 적용될 경우 초기에 임의 할당된 문장 점수에서 출발하여, 이전에 높은 점수를 부여 받은 많은 다른 문장들과 높은 유사도로 연결된 문장들에 이전보다

5) 그래프에서 한 노드의 차수(degree)는 그 노드와 연결된 다른 노드(들)의 개수이다.

더 높은 점수가 부여되도록, 문장 점수를 재귀적(recursively)으로 변경하는 방식으로 동작한다. 이 방법은 DUC[6]-2002, 2003, 2004에서 최상위 평가된 문서 요약 시스템과 대등한 성능을 보였다(Mihalcea, 2004 ; Erkan & Radev, 2004).

문장 그래프를 활용하는 문서 요약 기법 중 군집화 방법이 있다. 이는 먼저 입력 텍스트를 구성하는 문장 그래프에 대해 군집화(clustering) 알고리즘을 적용하여 유사한 문장들을 같은 그룹(군집, cluster)들로 묶은 다음, 각 군집에 대해 그 군집을 대표하는 문장을 하나씩 추출하여 요약문을 구성하는 방식을 취한다(Radev et al., 2004 ; Bossard et al., 2008). 하나의 문장을 그 문장 내의 단어들의 벡터로 표현할 때, 군집의 표현은 군집 내 문장 벡터들의 평균 벡터인 센트로이드(centroid)로 나타낼 수 있다. 따라서 군집의 대표 문장은 군집의 센트로이드와 가장 유사한 군집 내 문장으로 정할 수 있다. 문서 내 문장들의 각 군집은 문서 내에 빈출하는 주제들을 소주제별로 구분한 것에 해당한다.

3) 주제 지역성

초창기 문서 요약 연구자들은 주제 지역성을 실현하기 위해 입력 문서의 첫 단락, 마지막 단락, 그리고 각 단락의 첫 문장 및 마지막 문장에 주제성 가중치를 부여하는 방법을 사용해 왔다(Baxendale, 1958 ; Edmundson, 1969). 이들이 다룬 학술 문헌의 경우 주제 문장들이 전술한 바와 같은 위치적 제약을 갖는다는 가정은 요약문 추출에서 강인하고 효과적임이 보고되었다. 주제 지역성이 극대화되는 장르로 뉴스 기사가 있다. 뉴스 기사의 경우 문서 내 앞 부분(lead)에 위치한 문장들의 주제성이 가장 크고 이후 주제성이 점차 감소하도록 글이 작성되는 것이 일반적이다. 따라서 뉴스 기사 요약의 경우 기사의 첫 문장부터 순서대로 요약문의 정해진 크기만큼 문장들을 추출하는 방식이 있으며 이를 첫머리 기반 요약법(lead-based

6) DUC(Document Understanding Conferences)－http://duc.nist.gov/

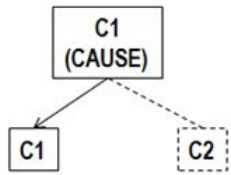
summarization)이라 한다. 이 방법은 현재까지도 단일 뉴스 기사 요약에서 가장 효과적인 기법 중 하나이다(Brandow et al., 1995 ; Nenkova, 2005).

이와 관련하여 문서 내에서 주제문의 위치가 장르에 따라 다를 수 있다는 가정에서 출발하여 장르 별 주제문의 최적 위치들에 해당하는 최적 위치 전략(Optimal Position Policy, OPP)을 자동 결정하는 방법이 시도되었다(Lin & Hovy, 1997). 이 방법은 사람이 작성한 요약문이 포함된 특정 장르의 문서 집합을 입력으로 받아 해당 장르에서 주제문의 선호되는 위치들을 단락번호, 단락 내 문장 번호의 형태로 제시한다. Ziff-Davis 코퍼스의 13,000 신문기사로부터 얻어진 OPP는 T, P2S1, P3S1, P4S1, P1S1, P2S2, {P3S2, P4S2, P5S1, P1S2}, P6S1의 순이었으며, Wall Street Journal의 경우 T, P1S1, P1S2의 순이었다(T는 제목 위치이고, PnSm은 n번째 단락 내 m번째 문장 위치를 의미한다). 최근 뉴스 기사 장르를 대상으로 하는 관련된 OPP 연구(Katragadda, 2009)가 있었으나 타 장르로의 적용 시도는 찾기 힘들다.

4) 담화 구조 주제 유추성

담화 구조로부터의 주제 확인은 입력 텍스트에 대해 만들어진 담화 구조 표현으로부터 주제문들을 추출하는 방식으로 동작한다. Marcu(1997)는 입력 문서를 분석하여 얻어진 수사구조트리(Rhetorical Structure Tree, RS-tree)의 상위 노드들에 해당하는 절들을 요약문으로 추출하는 시도를 하였다. 담화 구조 표현의 한 형태인 RS-tree는 텍스트의 모든 절(clause)들을 양보(concession), 인과(cause), 대조(contrast), 부연(elaboration) 등의 담화 관계(rhetorical relation)로 연결해 둔 것으로 담화 관계의 핵에 해당하는 절들이 트리의 상부에 위치한다. 수사구조이론(Rhetorical Structure Theory, RST)에 따르면 담화 관계로 연결되는 두 절은 핵(nucleus)과 위성(satellite)의 역할로 나뉘는데 이는 각각 주제 표현의 필수, 비필수적 요소에 해당한다. 예를 들어 한 문장 S가 'because'가 이끄는 원인절과 결과절로 이루어진 경우 원인절, 결과절은 각각 위성, 핵에 해당한다. S는 인과 관계를 표현하는 부모 노드 아래 결과절, 원인절을 두 자식 노드들로 갖는 트리의 형태로 표

현되며 부모 노드에는 두 자식 노드 중 핵에 해당하는 결과절이 대응된다 (<그림 1> 참조).



C1[James was happy] C2[because his son won a prize]

〈그림 1〉 RS-tree 예

Marcu는 텍스트의 수사구조 파싱(parsing)을 위해 미리 수집된 460개 담화 표지(discourse marker)를 중심적으로 활용하였다. 담화 구조 표현의 근간이 되는 담화 관계가 담화 표지로부터 유추될 수 있다는 점을 고려할 때 담화 표지는 텍스트의 주제 파악에 중요한 기여를 할 수 있다. Marcu가 사용한 담화 표지 중 "to conclude", "summing up", "on the one hand", "on the other hand", "it can be concluded that", "by contrast" 등의 단서구(cue phrase)들은 그 자체로 그들이 출현한 문장의 주제성 정도를 내포하고 있다고 볼 수 있다.

이와 관련하여 초기 문서 요약 연구에서는 학술 논문의 주제문들에서 특정한 단서구들이 동반 출현한다는 점에 착안하여 단서구 기반 요약(cue-based summarization)을 시도하였다. Edmundson(1969)은 미리 준비된 단서구 사전을 이용하여 'significantly', 'definitely', 'in particular' 등의 단서구가 출현한 문장의 주제성은 높이고 'hardly', 'impossible', 'unclear' 등이 출현한 문장의 주제성은 낮추는 방식으로 문장의 주제성을 계산하였다. Teufel과 Moens(1997)는 'argue', 'propose', 'develop', 'attempt', 'prove', 'show', 'conclude' 등을 포함하는 1,423개 단서구를 활용한 요약을 시도하였다. 단서구 기반 요약의 경우 텍스트에 출현한 단서구들의 모음으로 담화 구조 표현을 극히 단순화한 것으로 해석할 수 있다. 그러나 이 방법은 학술 문

헌과 같은 특정 분야 글에 제한적으로 적용되었으며 타 분야로의 일반화 시도는 찾기 힘들다.

5) 인간 요약 모방성

자동 요약을 위해 사람이 작성한 요약문의 특성을 활용할 수 있다. 글의 제목(title)은 글의 주제를 압축적으로 지시하는, 저자가 직접 작성한 지시적 요약문에 해당한다. 따라서 텍스트에 출현한 문장 중 텍스트의 제목과 가장 유사한 상위 n개 문장들을 요약문 후보로 추출할 수 있을 것이다. 이러한 제목 기반 요약 기법(title-based summarization)은 장, 절의 제목(heading)을 동시에 고려하기도 하였다(Edmundson, 1969).

제목 기반 요약 기법을 일반화시켜, 사람 요약문(예 : 초록)이 부착된 문서에서 사람 요약문과 문서의 나머지 문장들과의 관련성을 파악하여 문서의 요약을 체계적으로 시도하는 기법으로 기계학습(machine learning)이 사용되었다. 이 방법은 먼저 사람이 추출한 요약문이 부착된 문서들의 집합을 학습데이터로 사용하여 인간추출요약문장이 갖는 공통 자질들의 중요도를 학습모델로 만들어 둔다. 이후 새로운 입력 텍스트의 각 문장에 대해 학습된 자질의 출현 정도를 계산하여 해당 문장이 요약문일 가능성을 점수화하는 방식으로 동작한다.

최초의 기계학습 기반 요약 시스템에서는 naive-Bayes 분류기에 기반하여 인간추출요약문의 규칙성을 파악하는 5가지 자질로 문장 길이가 5단어 이상인지 여부, 문장이 단서구를 포함하고 있는지 여부, 문장의 위치가 단락의 처음 혹은 마지막인지 여부, 문장 내 주제 용어 포함 여부, 문장 내 대문자 용어(예 : UNESCO) 포함 여부를 사용하였다(Kupiec et al., 1995). Lin(1999)은 결정트리(decision tree) 기계학습법을 title similarity, TF IDF, OPP, query signature, IR signature, sentence length, numerical data, proper name, pronoun or adjective, weekday or month, quotation 등 총 18가지 자질의 학습에 적용하였다. 이 외에도 대표적으로 HMM(Conroy et al., 2001), CRF(Shen et al., 2007), Neural Network(Svore et al., 2007) 등의 기계학습 기법이 문서 요

약에 적용되었다.

6) 요약문 최소 중복성

추출 요약 시스템이 제시하는 발췌문들은 그들 간 정보 중복의 정도를 고려하지 않을 경우 같은 내용을 반복하는 서로 다른 문장들이 다수 포함될 위험이 있다. 군집법에 기반한 주제문 추출의 경우는 유사한 문장들이 모인 군집 별로 하나씩의 요약문 후보가 추출되므로 자연스럽게 요약문 중복성의 문제가 해결된다고 볼 수 있다. 그러나 그 외 요약 기법들은 요약문 후보 문장들 상호간 유사도를 비교하여 특정 임계치 이상의 유사도를 갖는 중복 문장들을 제거하는 별도의 절차를 수행할 필요가 있다.

질의 관련 다중 문서 요약(QFMDS)의 경우는 요약 시스템이 다중 문서 집합과 함께 질의문을 동시에 고려해야 한다. 즉 QFMDS의 요약문은 다중 문서의 내용을 대표하는 주제문을 찾아야 하는 동시에 그 주제문들이 질의문과 높은 관련성을 가져야 한다는 조건도 만족시켜야 한다. 이 두 가지 제약을 동시에 고려하는 대표적 QFMDS 요약 기법으로 최대 잔여 적합성(Maximal Marginal Relevance, MMR) 방법이 있다. MMR은 아래 수식이 보이는 바와 같이, 요약 대상이 되는 전체 문장들의 집합(S)으로부터, 질의(Q)와의 관련성(Sim1)이 높으면서 동시에 기 선택된 요약문장들(sj A)과의 관련도(Sim2)가 낮아야 하는 조건을 최대(argmax, max)로 만족시키는, 하나의 문장(s*)을 제거하여 요약 문장들의 집합(A)에 추가하는 절차를 반복하는 방식으로 동작한다(Carbonell & Goldstein, 1998).

$$s^* \equiv \arg max_{s_i \in S-A}\left[\lambda Sim_1(s_i, Q) - (1-\lambda)\max_{s_j \in A}\left(Sim_2(s_i, s_j)\right)\right]$$

2. 추상 요약

추상 요약은 입력 텍스트를 해석(interpretation)하여 내부 의미 표현을 만

들고 이를 압축한 형식의 요약 표현으로 변환(transformation)시킨 다음 자연어 생성 모듈을 통해 요약문을 생성(generation)하는 단계들을 거치는 것으로 정의된다(Sparck-Jones, 1999 ; Mani & Maybury, 1999). 그러나 문장을 구문적으로나 의미적으로 해석하는 기계적 프로그램들은 있으나 텍스트의 온전한 의미 표현을 분석해 내는 기계적 도구는 현재 존재하지 않는다. 다만 텍스트의 의미 표현을 담화 구조의 형식으로 제한하여 분석하는 해석기의 개발은 보고되고 있으나(Sumita et al., 1992 ; Marcu, 1997 ; duVerle & Prendinger, 2009), 추출식 요약에 적합한 의미 표현의 수준을 능가하기 힘들다.

따라서 기존 추상 요약 시스템들은 텍스트의 내부 의미 표현을 위해 특정 도메인에 대해 미리 정의해 둔 템플리트(template, 스키마, schema)를 주로 사용해 왔다(DeJong, 1978 ; McKeown & Radev, 1995). <그림 2>는 테러 기사 텍스트에 대한 의미 표현의 형태로, 테러 기사 글에서 언급되는 정보 항목의 유형(aspect)과 값을 Slot-Filler 쌍들의 집합으로 표현해 둔 템플리트의 한 예를 테러 기사 글의 예와 함께 보이고 있다. 이후 추상 요약 시스템은 정보 항목의 값이 채워진 템플리트를 자연어 생성 시스템에 입력으로 넣고 요약문을 생성하는 절차를 수행한다. 텍스트를 입력 받아 <그림 2>와 같은 템플리트의 Filler 값을 채우는 작업은 정보 추출(Information Extraction, IE) 시스템이 담당하므로 템플리트 기반 요약을 정보추출 기반 요약(IE-based summarization)이라고도 부른다.

정보추출 기반 요약은 템플리트가 작성된 특정 도메인에서 효과적일 수 있으나 같은 도메인 내에서도 템플리트의 Slot으로 지정된 정보 유형 외의 다른 새로운 것들을 인식하지 못하며 타 도메인으로의 적용 시 템플리트를 새롭게 작성해야 하는 단점이 있다. 그러나 다중 문서 요약의 경우 템플리트의 서로 다른 Slot들을 서로 다른 문서의 텍스트로부터 수집하는 과정에서 자연스럽게 다중 문서 요약이 수행되는 효과가 있다. 또한 일단 템플리트의 Filler가 채워지면 서로 다른 언어로의 자연어 생성이 가능하므로 교차 언어 문서 요약 시스템의 개발이 용이해 지는 장점[7]이 있다.

7) http://www.isi.edu/natural-language/teaching/cs544/apps3-summarization-new.pdf

19 March – A bomb went off this morning near a power tower in San Salvador leaving a large part of the city without energy, but no casualties have been reported. According to unofficial sources, the bomb – allegedly detonated by urban guerrilla commandos – blew up a power tower in the northwestern part of San Salvador at 0650 (1250 GMT).

Slot	Filler
Incident type	bombing
Date	March 19
Location	El Salvador: San Salvador (city)
Perpetrator	urban guerrilla commandos
Physical target	power tower
Human target	
Effect on physical target	destroyed
Effect on human target	no injury or death
Instrument	bomb

〈그림 2〉 테러 관련 텍스트와 템플리트 예(Grishman, 1997)

Ⅳ. 요약 평가 방법 및 요약 평가 대회

1. 요약 평가 방법

자동 요약을 포함하여 자연어정보처리시스템의 평가는 intrinsic, extrinsic 평가의 두 가지로 나뉜다(Sparck-Jones & Galliers, 1996). 자동 요약의 관점에서 intrinsic 평가는 시스템이 만들어 낸 자동 요약문(system summary, 시스템 요약문, 기계 요약문)이 본질적으로 사람이 작성한 요약문(human summary, 사람 요약문)과 일치하는 정도를 정보성(informativeness), 문법성, 응집성 등의 측면에서 측정하는 것으로 대부분 요약 시스템 평가는 이 방식으로 이루어진다. Extrinsic 평가는 요약이 아닌 다른 외부 작업(task)을 수행함에 있어, 요약 시스템이 출력한 요약문을 활용하는 것이 작업 수행 시간 및 완성도 측면에서 도움이 되는지를 측정하는 것이다.

Intrinsic 평가를 위해서는 요약 대상 문서에 대해 사람이, 주제문 추출 방식이 아닌, 직접 작성한 요약문이 필요하다. 그러나 자동 요약 커뮤니티

에서는 일반적으로 단일의 정답 요약문은 만들기 어렵다는 입장을 취한다(Mani et al., 1999 ; Hovy, 2005). 즉 서로 다른 사람은 서로 다르지만 올바른 요약문을 작성한다는 것이다. 따라서 공정한 평가를 위해서는 최소 2인 이상의 사람이 작성한 사람 요약문이 필요할 것이다. 이후 사람 요약문을 시스템 요약문과 비교하는 방식으로 수동(manual)과 자동(automatic) 기법을 고려할 수 있다. QFMDS 태스크에 대한 최신 수동 평가 지표로는, 기계 요약문이 사용자의 질의문을 만족하는 정도를 요약 내용과 언어적 표현 완성도 측면에서 1~5점 사이의 점수를 부여하는 Responsiveness 지표와, 요약 내용과 무관하게 기계 요약문의 문법성, 비중복성, 대용 표현 완전성, 포커스(focus), 구조, 일관성(coherence) 측면에서 1~5점 사이의 점수를 부여하는 Readability 지표가 있다(Dang & Owczarzak, 2008).

수동 평가는 소요되는 시간과 노력의 부담뿐 아니라 다른 사람에 의한 재반복(replication) 시 평가의 일관성이 보장될 수 있는가 하는 문제를 안고 있다. 보다 공정한 수동 평가를 위해 TAC 요약대회에서는 하나의 문서에 대해 서로 다른 시스템이 제출한 자동 요약문들을 같은 사람이 수동 평가하도록 하고 있다. 그러나 이 또한 이후 또 다른 시스템의 요약문에 대해 이전의 같은 사람이 수동 평가를 예전과 같은 일관성으로 재반복할 수 있는가 하는 문제는 해결되지 않는다. 이러한 수동 방식 요약 평가의 어려움은 기존 요약 기법의 개선이나 새로운 요약 기법 개발의 장애물이다. 다행히 최근 개발된 ROUGE-2, ROUGE-SU4, BE-HM 등의 자동 평가 지표들은 평균적으로 Responsiveness, Readability 등의 수동 평가 지표들과 높은 상관 관계를 보이는 것으로 보고되고 있다(Dang & Owczarzak, 2008).

ROUGE(Recall-Oriented Understudy of Gisting Evaluation)는 사람 요약문과 기계 요약문 사이에 단어 n-gram[8]이 겹치는 정도를 자동 계산하는 요약 평가 방법으로(Lin, 2004) 다양한 변형들이 있다. 이 중 ROUGE-2는 사람 요약문에 출현한 2-gram 중 기계 요약문에서 재현된 비율을 계산한 것이다.

8) 단어 n-gram은 문장에서 연속된 n개 단어들의 나열을 의미한다. “The king was happy”에서 1-gram(unigram)은 “The”, “king”, “was”, “happy”이고 2-gram(bigram)은 “The king”, “king was”, “was happy”이다.

ROUGE-SU4는 사람 요약문과 기계 요약문 사이에서 Unigram과 (최대 4 단어 거리 내에서 생성된) Skip-2gram[9]들의 재현율, 정확률의 조화평균을 계산한 것이다. 그러나 ROUGE는 문장을 일차원적 단어의 나열로 가정함으로써 이차원적 구문 구조를 고려하지 못하는 단점이 있다. 이 문제를 다루기 위해 제안된 BE(Basic Elements)는 사람 요약문과 기계 요약문 사이에 구문 단위의 중복 정도를 계산하는 방법이다(Hovy et al., 2005). BE-HM은 사람 요약문과 기계 요약문들을 각각 의존 문법에 따라 구문 분석한 결과에서 지배소(Head)－의존소(Modifier) 쌍[10]의 겹치는 정도를 요약 성능으로 계산한다.

2. 요약 평가 대회

최초의 대규모 자동 텍스트 요약 시스템들의 평가는 1998년 SUMMAC (TIPSTER Text Summarization Evaluation)에서 이루어졌으며 미국 정보 분석가들이 전형적으로 수행하는 검색, 분류의 두 유형 태스크에서의 자동 요약의 유용함을 평가(즉 extrinsic evaluation)하였다(Mani et al., 1999). 이 중 검색 태스크에서는 질의문에 대해 정보검색 시스템이 검색한 문서의 원문을 읽고 적합 문서를 찾는 작업이, 문서의 원문 대신 요약 시스템이 제시한 요약문만을 읽는 방식으로 변경했을 때, 작업 완료 시간과 작업 완성도 측면에서 어떻게 영향을 받는지를 평가한 것이다. 분류 태스크에서는 문서의 토픽을 5가지 중 하나로 분류하는 작업이 문서의 원문을 읽은 경우와 기계 요약문만 읽은 경우 얼마나 달라지는지를 평가한 것이다. 그 결과로

---

9) skip-2gram은 단어의 연속성 제약이 제거된 2-gram과 같다. “The king was happy”에서 skip-2gram은 “The king”, “The was”, “The happy”, “king was”, “king happy”, “was happy”이다.

10) “Two Libyans were indicted for the Lockerbie bombing in 1991”의 의존 구문 분석으로부터 추출된 지배소－의존소 쌍들은, 의존 관계를 같이 표현하면, [Libyans|two|nn], [indicted|Libyans|obj], [bombing|Lockerbie|nn], [indicted|bombing|for], [bombing|1991|in]과 같다(Hovy et al., 2005).

요약문에 기초한 검색, 분류 작업은 원문에 기초한 작업에 비해 각각 5%, 14%의 낮은 성능 저하를 보이면서 작업 완료 시간 측면에서 각각 50%, 40%의 시간을 줄인 것으로 보고되어 요약의 유용함과 향후 요약 연구의 필요성이 입증되었다.

이후 2001~2007년까지 진행된 DUC(Document Understanding Conferences)에서는 뉴스 기사를 대상으로 하여 단일문서요약(SDS), 다중문서요약(MDS)과 관련된 태스크에 대해 기계 요약문의 본질적 평가(즉 intrinsic evaluation)를 중점적으로 수행하였다. 2001~2002년 동안 진행된 Generic SDS는 단일 문서에 대해 100 단어 길이의 generic summary를 자동 생성하는 태스크였다. 2001~2004년 동안 진행된 Generic MDS는 다중 문서 집합(약 10개 문서)에 대해 50, 100, 200, 400 단어 길이의 generic summary를 자동 생성하는 태스크였다. 2005~2007년 동안 진행된 Focused MDS(QFMDS와 같다)는 질의와 관련된 다중 문서집합(약 25개 문서)에 대해 질의와 관련된 250 단어 길이의 요약문을 작성하는 태스크였다.

Generic SDS 태스크의 결과(Nenkova, 2005)로 어떤 참가 시스템들도 문서의 첫 100단어를 요약문으로 추출하는 lead-based 요약문(baseline)보다 우수한 요약문을 생성하지 못하였으나 대부분의 사람 요약문들은 baseline보다 우수했으므로 향후 이 태스크의 성능 향상이 가능함을 보였다. 그러나 일부 사람 요약문(9명 중 2명)은 최상위 기계 요약문보다 성능이 낮았다. 이는 대부분 사람들은 기계보다 우수한 요약문을 작성하나 일부 사람들의 요약 전략은 그렇지 않음을 시사하는 것으로, 요약 평가와 관련하여 사람 요약문 작성을 위해 다수의 인원이 필요함을 의미한다. 2002년 이후 Generic SDS는 DUC, TAC에서 다루어지지 않았다.

DUC의 뒤를 이어 2008년 이후 계속되고 있는 TAC[11](Text Analysis Conference)에서는 뉴스 기사를 대상으로 Focused MDS와 그 변형 태스크들을 평가해 왔다. Update Focused MDS(줄여서 Update summarization)(2008~2009년)는 사용자가 질의와 관련된 이전 10개 다중 문서를 이미 읽었다고 가정

11) http://www.nist.gov/tac/

하고 이후 같은 질의와 관련되어 새롭게 출현한 10개 문서에 대한 요약문(update summary)을 작성하는 태스크로, 이전 10개 다중 문서에 대한 요약문(initial summary) 작성도 함께 평가된다. 2010년 8월 현재 TAC에서는 자연재해, 테러, 멸종 위기 자원 등 특정 카테고리에 속하는 질의와 관련된 10개 다중 문서에 대해 질의가 속하는 카테고리와 관련하여 미리 지정된 여러 속성(aspects)[12] 정보와 관련된 100 단어 요약문을 작성하는 Guided Focused MDS(줄여서 Guided summarization) 태스크를 진행하고 있다. 이와 동시에 Guided summarization에 대한 Update summarization 태스크도 진행 중이다.

<표 1>은 최근 2008년 TAC의 Update summarization 태스크의 평가 수치들을 정리한 것이다. Baseline은 10개 문서 중 가장 최근 문서의 첫 100 단어 요약문을 추출한 것이다. Responsiveness, Readability 등 수동 평가에서 기계 요약문은 사람 요약문에 크게 못 미치고 있다. 특별히 한 문서 내 연속된 문장들의 나열인 baseline 요약문은, 내용을 고려하는 Responsiveness 측면에서 기계 요약문에 뒤지고 있으나, 내용을 고려하지 않고 문법성, 응집성 등을 판단하는 Readability에서는 기계 요약문을 능가하고 있다. 사람 요약문들과 기계 요약문들의 평균적 성능에 있어, 자동 평가 지표는, 사람 요약문이 기계 요약문을 능가한다는, 수동 평가 지표의 결과와 높은 상관성이 있는 것으로 보고하였다(Dang & Owczarzak, 2008). 그러나 기계 요약문의 평균적 성능이 아닌 최상위 성능을 고려할 경우 자동 평가 지표들은 기계 요약문이 최하위 사람 요약문과 대등하다는, 보다 정확한 수동 평가의 결과와는 다른, 결과를 보이고 있다. 이는 현재 문서 요약에서 사용되는 최신 자동 평가 지표의 개선이 필요함을 시사하는 측면 중 하나이다.

12) 자연재해(natural disaster) 카테고리의 경우 what happened ; date ; location ; reasons for accident/disaster ; casualties ; damages ; rescue efforts 등의 속성(aspects)을 미리 정의하고 있다(http://www.nist.gov/tac/2010/Summarization/).

〈표 1〉 TAC-2008 (Update summarization) 수동, 자동 요약 평가 결과(Dang & Owczarzak, 2008)

	Metric	Initial summary			Update summary		
		Human	Baseline	Machine	Human	Baseline	Machine
Manual	Responsiveness	4.41 ~ 4.79	2.29	1.29 ~ 2.79	4.29 ~ 4.87	1.85	1.10 ~ 2.60
	Readability	4.62 ~ 4.91	**3.25**	1.37 ~ 2.93	4.58 ~ 4.95	**3.41**	1.18 ~ 3.20
	Pyramid evaluation	0.52 ~ 0.84	0.18	0.08 ~ 0.35	0.49 ~ 0.76	0.14	0.03 ~ 0.33
Automatic	ROUGE-2	**0.108** ~ 0.13	0.05	0.03 ~ 0.111	**0.106** ~ 0.13	0.05	0.01 ~ 0.101
	ROUGE-SU4	**0.140** ~ 0.17	0.09	0.07 ~ 0.142	**0.136** ~ 0.16	0.09	0.04 ~ 0.136
	BE-HM	**0.067** ~ 0.09	0.03	0.01 ~ 0.063	**0.073** ~ 0.10	0.03	0.01 ~ 0.075

Baseline means the first few sentences of the most recent document

Ⅴ. 결론 및 향후 과제

다양한 출처로부터의 대규모 정보들을 읽고 정리하는 시간을 단축해야 하는 인간의 요구는 텍스트 자동 요약 연구를 이끌어 온 내재된 힘이었다. 1950년대 학술 논문의 요약에서 시작하여 현재까지 뉴스 기사, 전자메일(Corston-Oliver et al., 2004), 회의 기록(Murray et al., 2005), 강의(Nobata et al., 2003), 구어체 대화(Zechner, 2001) 등 다양한 장르 텍스트의 요약이 시도되었다.[13] 특히 뉴스 기사의 요약은 단일, 다중 문서 요약을 포함하여 지난 10년간 DUC, TAC에서 수행된 대규모 요약 시스템 평가의 지속적 주제였다. 이 논문에서는 추출 방식을 중심으로 텍스트 자동 요약 기술의 현황을 제시하였으며 요약 평가 방법과 대규모 자동 요약 대회에 대한 개괄을 기술하였다.

그간의 텍스트 자동 요약 기술은 텍스트의 표층에서 발견되는 어휘의 통계적 특성에 주로 의존하는 추출 요약 기법이 대부분이었다. 추출 요약의 경우 주제 확인을 위해 용어 집합, 어휘/문장 그래프와 같은 문서 표현을 사용하고 있으나 이들은 주로 텍스트 표층 수준의 어휘에 기초하여

13) 향후 새로운 장르(예 : 비유적 표현이 풍부한 문학작품 등) 텍스트에 대한 자동 요약이 시도될 수 있겠으나 뉴스 기사와 같은 사실적 텍스트가 아닌 경우 사람의 요약에 근접하는 실용적 요약 시스템은 가까운 미래에 실현되기 힘들 것이다.

만들어지고 있다. 어휘의 의미나 개념에 기초한 보다 심층 수준의 문서 표현을 위해서는 어의 중의성 해소, 대용어 참조 해소, 어휘개념망/온톨로지 기반 개념 관계 및 유사도 측정 기법 등이 전제되어야 한다. 텍스트의 온전한 의미 표현이 전제되어야 하는 추상 요약 기술 역시 특정 분야 글이 다루는 핵심 소재들을 미리 정의된 템플리트 형식으로 이해하는 수준에 그치고 있다.

추출 요약 기법이 모든 유형의 텍스트에 대해 고속의 강인한 요약 결과를 만들 수 있으나 인간의 요약 수준과는 여전히 큰 차이가 있다. 현재의 자연어처리기술 수준을 감안할 때 텍스트의 온전한 의미 표현을 만드는 자동 기법의 출현은 요원하다. 단기적으로 두 문장 혹은 한 단락의 내용을 압축하여 하나의 새로운 문장을 생성하는 언어학적 기법의 개발이 현재 추출 중심 요약의 한계를 극복하는 데 도움이 될 것이다. 텍스트에 대해 문장 단위의 담화 관계를 인식해 둔 담화 구조 표현으로부터 요약문을 추출하는 기법(Marcu, 1997) 또한 현재 추출식 요약의 한계를 극복하는 기대되는 시도 중 하나이다. 이와 관련된 한 연구에서는 담화 관계가 부착된 말뭉치로부터의 기계학습을 통해 93%의 담화 표지 분류 성능을 보고하였다(Pitler et al., 2008). 또한 향후에도 DUC, TAC 등과 같은, 요약 기법에 대한 대규모 평가를 지속적으로 유지하면서 현재 자동 요약 평가 지표가 갖는 문제 해결을 시도하는 새로운 자동 평가 기법의 개발(Lin et al., 2006)이 필요하다.

대부분의 자동 요약 기법들은 영어를 대상으로 연구되었으나 요약 기법 자체가 언어에 종속적인 것은 아니다. 즉 한국어를 대상으로 하는 자동 문서 요약 연구들도 요약 기법 자체와 관련하여서는 이 글에서 살핀 요약 기법의 범주를 크게 벗어나지 않는다. 다만 자동 요약 시스템의 수준은 특정 언어의 자연어처리 도구의 정교함에 의존하므로 한국어 자동 문서 요약 연구의 발전을 위해서는 향후 한국어 텍스트의 이해 및 분석 연구의 진보가 전제되어야 할 것이다.*

* 본 논문은 2010. 10. 30. 투고되었으며, 2010. 11. 5. 심사가 시작되어 2010. 11. 29. 심사가 종료되었음.

▣ 참고문헌

Barzilay, R. & Elhadad, M.(1997), Using lexical chains for text summarization, In Proceedings of the ACL/EACL Workshop on Intelligent Scalable Text Summarization.

Baxendale, P.(1958), Machine-made index for technical literature-an experiment, IBM Journal of Research and Development, 2 : 354-361.

Bossard, A., Genereux, M. & Poibeau, T.(2008), Description of the lipn systems at TAC 2008 : summarizing information and opinions, In Proceedings of the Text Analysis Conference.

Brandow, R., Mitze, K. & Rau, L.(1995), Automatic condensation of electronic publications by sentence selection, Information Processing and management, 31(5) : 675-686.

Brin, S. & Page, L.(1998), The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 30 : 1-7.

Carbonell, J. & Goldstein, J.(1998), The use of MMR and diversity-based reranking for reordering documents and producing summaries, In Proceedings of SIGIR-1998, 335-336.

Conroy, J. & O'leary, D.(2001), Text summarization via hidden markov models, In Proceedings of SIGIR, 406-407.

Corston-Oliver, S., Ringger, E., Gamon, M. & Campbell, R.(2004), Task-focused summarisation of email, In Proceedings of ACL, 43-50.

Dang, H. T. & Owczarzak, K.(2008), Overview of the TAC 2008 update summarization task, In Proceedings of the Text Analysis Conference, 10-23.

DeJong, G.(1978), Fast Skimming of News Stories : The FRUMP System, Ph.D. Dissertation, Yale University.

duVerle, D. & Prendinger, H.(2009), A novel discourse parser based on support vector machine classification, In Proceedings of ACL, 665-673.

Edmundson, H.(1969), New methods in automatic extracting, Journal of the ACM, 16(2) : 264-285.

Erkan, G. & Radev, D.(2004), LexRank : graph-based centrality as salience in text summarisation, Journal of Artificial Intelligence Research, 22 : 457-479.

Grishman, R.(1997), Information extraction : techniques and challenges, In Pazienza, M. (ed), Information Extraction, Springer, Berlin.

Halliday, M. & Hasan, R.(1976), Cohesion in English, London : Longman.

Hovy, E.(2005), Automated text summarization, In Mitkov, R. (ed), The Oxford Handbook of Computational Linguistics, Oxford : Oxford University Press.

Hovy, E., Lin, C. & Zhou, L.(2005), Evaluating duc 2005 using basic elements, In Proceedings of Document Understanding Conference(DUC).

Katragadda, R., Pingali, P. & Varma, V.(2009), Sentence position revisited : a robust light-weight update summarization baseline algorithm, In Proceedings of the International Workshop on Cross Lingual Information Access, 46-52.

Kleinberg, J.(1999), Authoritative sources in a hyperlinked environment, Journal of the ACM, 46(5) : 604-632.

Kupiec, J., Pedersen, J. & Chen, F.(1995), A trainable document summariser, In Proceedings of SIGIR, 68-73.

Lin, C. & Hovy, E.(1997), Identifying Topics by Position, In Proceedings of ANLP, 283-290.

Lin, C.(1999), Training a selection function for extraction, In Proceedings of CIKM, 1-8.

Lin, C.(2004), Rouge : a package for automatic evaluation of summaries, In Proceedings of ACL, 74-81.

Lin, C., Cao, G., Gao, J. & Nie, J.(2006), An information-theoretic approach to automatic evaluation of summaries, In Proceedings of HLT-NAACL, 463-470.

Luhn, H.(1958), The automatic creation of literature abstracts, IBM Journal of Research and Development, 2 : 159-165.

Mani, I. & Maybury, M.(1999), Advances in automatic text summarisation, Cambridge, M A : MIT Press.

Mani, I., Firmin, T., House, D., Klein, G., Sundheim, B. & Hirschman, L.(1999), The TIPSTER SUMMAC text summarization evaluation, In Proceedings of EACL, 77-85.

Marcu, D.(1997), The rhetorical parsing, summarization, and generation of natural language texts, Ph.D. thesis, University of Toronto.

McKeown, K. & Radev, D.(1995), Generating summaries of multiple news articles, In Proceedings of SIGIR, 74-82.

Mihalcea, R.(2004), Graph-based ranking algorithms for sentence extraction, applied to text summarization, In Proceedings of ACL.

Morris, A., Kasper, G. & Adams, D.(1992), The effects and limitations of automated text condensing on reading comprehension performance, Information Systems Research, 3(1) : 17-35.

Murray, G., Renals, S. & Carletta, J.(2005), Extractive summarisation of meeting recordings, In Proceedings of ACL.

Nenkova, A.(2005), Automatic text summarization of newswire : Lessons learned from the document understanding conference, In Proceedings of AAAI.

Nobata, C., Sekine, S. & Isahara, H.(2003), Evaluation of features for sentence extraction on different types of corpora, In Proceedings of ACL.

Pitler, E., Raghupathy, M., Mehta, H., Nenkova, A., Lee, A. & Joshi, A.(2008), Easily identifiable discourse relations, In Proceedings of COLING.

Radev, D., Jing, H. & Budzikowska, M.(2000), Centroid-based summarisation of multiple documents : sentence extraction, utilitybased evaluation, and user studies, In Proceedings of ANLP/NAACL, 21-30.

Salton, G., Singhal, A., Mitra, M. & Buckley, C.(1997), Automatic text structuring and summarization, Information Processing & Management, 33(2) : 193-207.

Shen, D., Sun, J., Li, H., Yang, Q. & Chen, Z.(2007), Document summarization using Conditional Random Fields, In Proceedings of IJCAI, 1805-1813.

Sparck-Jones, K. & Galliers, J. R.(1996), Evaluating natural language processing systems, Berlin : Springer.

Sparck-Jones, K.(1999), Automatic summarising : Factors and directions, In Mani, I. & Maybury, M. (ed), Advances in automatic text summarisation. Cambridge, MA : MIT Press.

Sumita, K., Ono, K., Chino, T., Ukita, T. & Amano, S.(1992), A discourse structure analyzer for Japanese text, In Proceedings of the International Conference on Fifth Generation Computer Systems, volume 2, 1133-1140.

Svore, K., Vanderwende, L. & Burges, C.(2007), Enhancing single-document summarization by combining RankNet and third-party sources, In Proceedings of EMNLP-CoNLL, 448-457.

Teufel, S. & Moens, M.(1997), Sentence extraction as a classification task, In Proceedings of ACL, 58-65.

Zajic, D. Dorr, B., Lin, J. & Schwartz, R.(2007), Multi-candidate reduction : Sentence compression as a tool for document summarization tasks, Information Processing and Management, 43 : 1549-1570.

Zechner, K.(2001), Automatic generation of concise summaries of spoken dialogues in unrestricted domains, In Proceedings of SIGIR, 199-207.

<초록>

문서 자동 요약의 현황과 과제

강인수

인간이 다루어야 할 정보가 기하급수적으로 증가하는 문제를 다루기 위해 전산언어학 및 자연어처리 커뮤니티에서는 문서 요약의 자동화 기법이 연구되고 있다. 1950년대부터 시작된 자동 문서 요약 연구는 여러 유형의 문서를 다루면서 단일/다중 문서 요약, 질의 관련 다중 문서 요약 등의 다양한 태스크에 적용하기 위한 추출 및 추상 방식 요약 기법을 시도해 왔다. 이 논문은 추출 방식을 중심으로 텍스트 자동 요약 기술의 현황을 제시하고 요약 평가 방법과 대규모 자동 요약 대회에 대한 개괄 및 향후 과제에 대해 기술한다.

【핵심어】 문서 자동 요약, 추출 요약, 추상 요약, 다중 문서 요약, 요약 평가

<Abstract>

Automated Text Summarization : a Survey

Kang, In-su

Information that human should read grows exponentially. To deal with this problem, computational linguistics and natural language processing communities have attempted to automate summarizing text. Since its start in 1950's, automated text summarization has handled single-/multi-document summarization using extracting and abstracting techniques, and nowadays specialized its tasks to query-focused multi-document summarization. This paper gives the current state of automatic text summarization techniques focusing on robust, practical extraction-based methods, and describes evaluation methodologies and large-scale summarization evaluation conferences. Finally, future issues are discussed.

【Key words】 Automated text summarization, Extractive summarization, Abstractive summarization, Multi-document summarization, Summarization evaluation