

단일문서와 복수문서 자동요약의 특성에 따른 기능 분석

Analysis on Automatic Summarization Functions of the Single Document and the Multi Documents

최상희, 연세대학교 대학원 문헌정보학과

Choi, Sang-hee, Dept. of Lib. & Info. Sci., Graduate School of Yonsei University

요약은 원문의 주제를 파악하여 원문의 축약판을 만들어 이용자에게 제공하는 중요한 정보 생산 과정이다. 최근 이용자에게 제공되는 정보량이 급증하면서 자동 요약에 대한 필요성이 더욱 증가하고 있으며 단일문서의 내용을 파악하는 도구로써 활용되던 요약이 문서집합의 내용을 파악하는 도구 및 새로운 정보생성 수단으로 그 기능을 넓혀가고 있다. 본 논고에서는 자동요약의 기본 개념과 요약대상의 문서 수에 따른 요약 특성 및 기능을 고찰하였다.

1 서 론

웹의 등장 및 정보시스템의 발전으로 이용가능한 정보량이 급증하게 되자, 검색결과로 나오는 정보량이 이용자의 물리적인 시간과 능력을 초과하는 경우가 빈번하게 발생하게 되었다. 그 결과, 정보의 내용을 간략하게 전달하는 자동요약은 이에 대한 해결방안으로 많은 연구자들의 주목을 받게 되었고 단일문서 요약이 대부분의 연구과제를 차지했던 자동요약연구는 복수문서나 멀티미디어 등, 여러 분야로 확장되기 시작했다.

최근 들어 복수문서의 자동요약은 요약의 기본 역할인 원정보원의 내용을 간략하

게 표현하는 기능에서부터 새로운 정보원을 생성하는 방향으로 발전하여 자동요약 연구의 새로운 방향성을 제시하고 있다. 본 논고에서는 자동요약의 과정과 특성, 유형을 분석하고, 단일문서와 복수문서의 특성을 비교한 후 두 요약방식의 적용 방안 및 기능에 대하여 고찰하고자 한다.

2 자동요약 유형 및 과정

2.1 요약의 유형

요약을 나누는 전통적인 기준은 원정보원의 가공수준과 기능, 언어 등이 있었지만 최근 자동요약 연구에서는 요약대상의 수, 이용자 지향여부, 매체, 갱신여부와 같은

새로운 기준들을 추가하여 요약의 유형을 규정하고 있다.

1) 정보내용의 가공수준에 따라 - 추록(extraction)과 초록(abstraction)

추록에서 핵심과제는 주제를 가장 잘 대표할 수 있는 정보를 원정보원에 있는 그대로 잘 선정(selection)하는 것이다. 반면 초록은 이표기로 분산되는 동일 개념을 파악하거나 상위개념이나 하위개념을 인식할 수 있는 지식베이스를 활용하여 주제를 잘 대변할 수 있는 요소를 잘 분석하여 새롭게 생성(generation)하는 것이다.

2) 요약의 기능에 따라 - 지시적 요약, 통보적 요약, 비평적 요약

이용자가 원정보원의 선택여부를 결정할 수 있도록 판단에 필요한 정보를 제공하는 지시적 요약(indicative summarization), 이용자가 원정보원을 보지 않아도 될 정도로 원정보원의 정보를 충분히 수록한 통보적 요약(informative summarization), 요약자의 견해가 개입되어 원정보원을 평가한 정보가 요약에 결합되는 비평적 요약(evaluative summarization)이 있다.

3) 원정보원의 언어에 따라 - 원어 요약과 번역 요약

단일언어로 쓰여진 정보원을 요약하는 경우는 언어문제를 크게 고려하지 않고 원어로 요약을 생성한다. 그러나 복수언어로

기술된 정보원을 요약할 때 요약은 한 언어로 기술하고자 하여 다른 언어로 번역하여 요약하는 경우가 있다.

4) 요약대상의 수에 따라 - 단일문서 요약과 복수문서 요약

단일문서의 요약은 한 문서의 내용을 요약하는 경우에 해당하지만 복수문서의 요약은 보통 같은 주제를 다루고 있는 문서 집합이나 연속적으로 입력되는 문서들을 요약하는 경우가 일반적이다.

5) 이용자 지향여부에 따라 - 문서 중심의 요약과 이용자 중심의 요약

문서 중심의 요약(generic summarization)은 원문서가 수록하고 있는 가장 중요한 주제를 기준으로 요약을 생성한 것으로 원문서의 축소판 내지는 대응물이 되는 것이다. 이용자 중심의 요약(user-focused summarization)은 이용자가 알고 싶은 주제나 정보 수준에 따라 요약을 생성해 주는 것이다. 따라서 이용자 중심의 요약은 다양한 이용자의 정보요구대로 요약시스템이 파악한 주제 외에 다른 주제로 생성 가능하다.

6) 요약대상의 매체에 따라 - 텍스트 요약과 멀티미디어 요약

최근 들어 요약의 대상은 음향이나 영상 자료로 확대되어 가고 있는 추세이다. 음향이나 영상자료 요약에 대한 접근은 음향이나 영상자료에 붙어있는 캡션자료를 활용하는 것에서부터 패턴인식을 적용하여 주

요 장면을 포착하는 것까지 다양하게 시도 되고 있다.

7) 정적, 동적 여부에 따라 - 정적 요약과 동적 요약

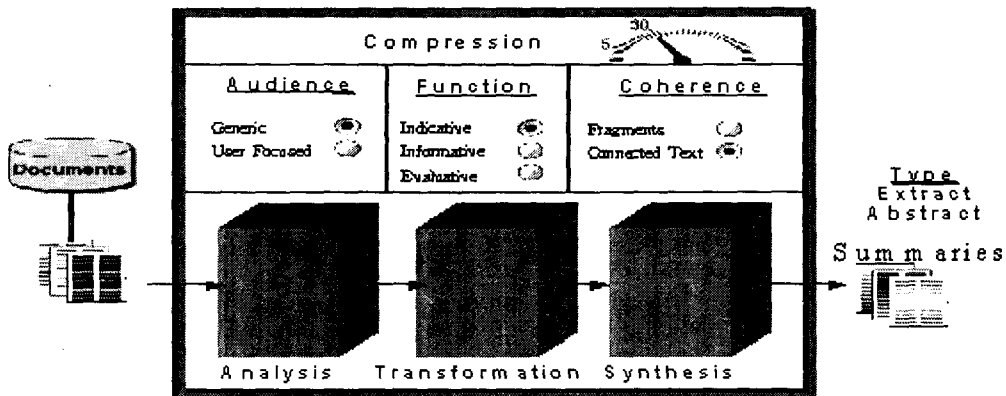
정적 요약은 한번 요약을 생성한 후 다시 갱신하지 않는 요약을 말하는 것이고, 동적 요약은 새로운 정보가 추가될 때 이를 추적하여 요약을 갱신하는 것을 말한다. 보통 정적 요약은 단일문서 요약에 적용되는 경우가 많고 동적 요약은 복수문서 요약에 적용되는 경우가 많다.

2.2 요약 과정 분석 및 제시 - Mani의 요약이론 중심으로

일반적으로 자동요약과정을 정리한 것 중 가장 많이 이용되는 것은 Mani(2001)의 개념도이다(그림 1 참조). Mani는 요약의 과정을 분석(analysis), 변환(transformation,

refinement), 합성(synthesis)으로 나누고 있다. 그러나 실제 원문에 나와 있는 것을 그대로 사용하는 추록(extraction)에서는 분석단계에서 가장 주제연관성이 높은 정보요소를 분석하여 원정보원에 있는 그대로 제공하는 것으로 끝나게 되어 실제로 변환이나 합성단계로 가지 않는 경우가 많다. 예를 들면 문장기반 추록시스템은 원정보원에서 주제를 나타내는 문장을 뽑아 내어 문장이 나타난 순이나 문서주제와 유사도 순으로 재배열하는 수준으로 요약을 생성하는 것이 일반적이므로 문장을 재구성하거나 여러 문장을 합성, 분리하는 과정을 포함하고 있지 않다. 반면 초록은 원정보원에 포함되어 있는 개념을 요약하는 것을 목적으로 하기 때문에 원정보원에서 추출한 용어나 문장을 그대로 사용하는 것보다 더 수준높은 처리과정을 필요로 한다.

초록(abstraction)은 Mani의 개념도에서 분석이후 단계인 변환과 합성이 가장 중요



〈그림 1〉 Mani의 요약개념도 (Mani 2001)

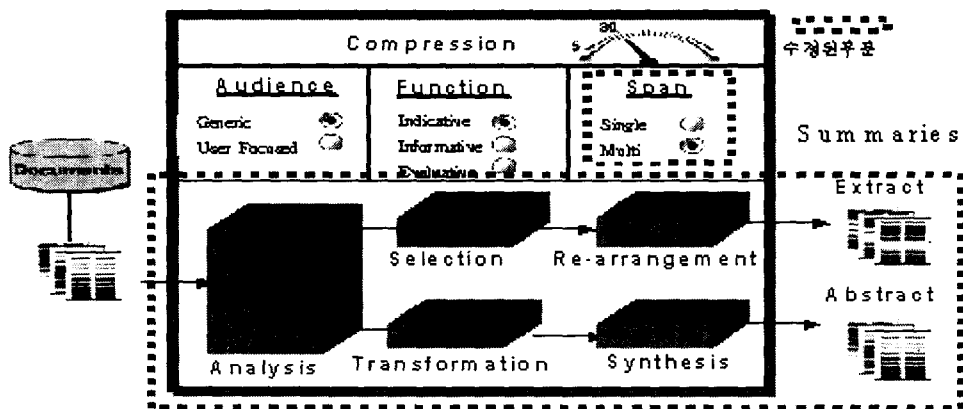
한 과정이 된다. 추출된 정보요소를 가지고, 새로운 문장이나 단락을 생성하여 요약 을 만드는 과정은 변환과 합성단계에서 이루어지기 때문이다. 이런 특성이 추록의 과정과 초록의 과정을 구분짓는 차이점이고 두 과정은 중점을 각기 다른 단계에 두고 있다고 할 수 있다. 초록시스템은 상대적으로 추록시스템보다 잘 구축된 지식기반시스템과 의미론적인 분석이나 구문적 분석 등 복잡한 정보처리 단계를 요구한다. 따라서 두 과정을 분리하지 않은 Mani의 개념도는 일반적인 요약의 개념도라고 여겨지기보다는 초록의 개념도에 적합하다고 할 수 있다.

각각의 특성을 고려해보면 원문에 있는 정보요소를 그대로 이용하는 추록에서는 변환의 개념보다 선정(selection)의 개념이 더 적절하다. 원문의 요소들을 분석한 후 가장 요약에 적합한 것을 선택하는 과정이 자동추록에서의 핵심이 되기 때문이다. 따

라서 분석과정을 좀더 세분하는 것이 자동 추록을 이해하는데 더 효율적일 것이다.

추록에서 특정 정보원의 내용을 표현하는 요소를 분석하기 위해서는 먼저 내용을 표현하는 요소 단위를 선정하는 것이 초록보다 더 중요한 문제가 될 수 있다. 선정된 단위가 그대로 요약에 적용되기 때문이다. 현재 요약연구에서는 문장 단위로 정보를 추출하는 경우가 가장 많다. 문장은 정보원의 주제를 나타내는 주요 용어를 포함하고 있으면서 그 주요 용어간의 관계를 담을 수 있는 가장 적절한 단위이기 때문이다. 그러나 최근에는 문장 외에도 단락 (passage)를 추출하여 요약의 가독성을 높이고 맥락적인 측면에서 주제정보를 보충하는 경우도 있다.

초록의 마지막 단계인 합성 개념은 추록에서 재배치(re-arrangement)라는 개념으로 대치되는 것이 더 적절하다. 추록은 추출된 정보를 그대로 사용하기 때문에 추출



〈그림 2〉 수정된 Mani의 요약 개념도

한 정보끼리 새로이 결합되는 경우가 드물기 때문이다. 그러므로 합성된다기 보다는 중요한 정보 순이나 요약의 주제적 맥락을 고려하여 다시 재배치하는 것이다.

위와 같은 관점에서 개념도를 다시 조정하면 <그림 2>와 같다. 수정된 개념도에서 요약과정 외에 추가로 변경한 사항은 원래의 개념도에 있었던 변수 중 “coherence” 부분을 문서 수(Span)로 변경한 것이다. “coherence”의 fragments는 요약이 용어리스트나 구, 고유명사 등으로 추출되는 경우인데 실제 요약에서는 결과물이 fragments로 되는 경우가 거의 없고 오히려 색인이나 질의응답에서 활발히 적용되는 변수라고 할 수 있으므로 실제 자동요약에는 적합한 변수라고 할 수 없다. 반면 문서의 수는 요약하는 대상의 범위를 조절함에 따라 요약 과정이 추가되거나 확장되는 등 다양한 영향을 미치므로 자동요약시스템에 직접변수 역할을 할 수 있다.

수정된 개념도는 결과물에 초점을 맞춘 것으로 요약의 결과물이 추록이나 초록이나에 따라 생성과정이 2가지로 나뉜다고 본 것이다(그림 2 참조). 추록인 경우에는 선정(selection)과 재배치(re-arrangement)라는 개념에 중점을 두어 요약과정을 규정하는 것이고, 초록인 경우에는 변환

(transformation)과 합성(synthesis)을 강조한 것이다. 초록과정에서도 선정과 재배치 과정이 분석과 변환과정에 포함되어 있을 수 있으나, 최종 산출물인 요약의 품질을 결정하는 영향력 측면에서 보면 추록에서처럼 의존도가 높다고 할 수 없다. 따라서 추록에서의 선정과 재배치 과정은 초록과정처럼 분석에 포함되어 있는 것보다는 훨씬 다양한 접근 방식이 적용된다고 할 수 있어 분리하는 것이 바람직하다.

3 단일문서 요약의 특성

단일문서의 요약을 보고자 하는 이용자의 목적은 원정보원을 읽지 않고 그 정보원의 내용을 파악하고자 하는 것이 가장 보편적이다. 따라서 단일문서의 목표는 원정보원의 정보를 최대한 효율적으로 표현하는 것이다.

■ 중복성(redundancy)

주제적인 측면을 보자면 단일문서 요약에서 특정 정보원의 주제를 파악하는데 있어 가장 중요하게 여기는 것은 중복성(redundancy) 또는 유사성(similarity)이다. 한 정보원에서 여러 번 반복되어 나타나는

<표 1> 단일문서와 복수문서 특성 비교

요약 환경	특성	주제의 중복성	주제의 고유성	원문서 길이	문서길이 차이	압축률 정도/적용방식	문서구조 활용도	번역기능 적용
단일문서		중요	고려안함	제한	보정안함	중간/획일적	높음	낮음
복수문서		중요	중요	제한안함	보정	높음/가변적	중간	중간

개념은 그 정보원의 주제가 될 가능성이 높다는 것이다. 예를 들면 한 문헌내의 문장간의 유사도를 측정했을 경우, 문장 A가 많은 문장들과 문장간 중복도나 유사도가 높게 나온다면 A는 그 문헌전체에서 여러 번 반복해서 나오는 개념을 포함하고 있는 것이고, 그 문헌의 주제를 담고 있다고 추정할 수 있다. 그러므로 문장A는 그 문서의 요약에 포함될 우선순위가 높아지게 된다.

이와 같은 논리로 가장 유사성이 높거나 중복도가 높은 문장들을 추출하면 그 정보원의 주제를 나타내는 요약이 된다는 것이다. 특히, 단일문서는 여러 가지 주제가 혼합되어 있거나 다양한 시각이 포함되어 있기 보다는 제한된 수의 주제와 시각으로 기술된 사례가 일반적이어서, 한 두 번 언급한 것은 실제 중요하지 않은 정보일 가능성이 높다. 그러므로 단일문서 요약에서는 다양한 주제를 파악하는 개념인 주제의 고유성(uniqueness) 또는 상이성(difference)을 거의 고려하지 않는다.

■ 원정보의 길이

요약을 보고자 하는 이용자는 원정보원의 내용을 파악하는데 더 적은 시간과 노력을 들이려고 한다. 따라서 원정보원을 보는데 드는 시간은 요약을 보는 시간보다 훨씬 많이 걸려야 한다는 전제가 생긴다. 그러므로 원정보원은 반드시 요약을 해서 효과를 볼 수 있을 정도의 길이가 되어야 한다. 너무 짧은 문서는 요약을 보는 것과 원문을 보는 것이 크게 차이가 없기 때문

에 요약의 대상으로 적합하지 않다. 복수문서의 요약에서는 요약의 원정보원이 한 문서가 아니라 여러 문서에 해당하는 것이므로 비록 문서의 길이가 짧더라도 여러 문서를 합치면 일정 수준 이상의 길이가 된다. 그러므로 복수문서 요약에서는 원정보원의 길이가 크게 문제 되지 않으나 문서 하나를 대상으로 하는 단일문서는 문서의 길이가 요약의 적용 여부를 결정하는 중요한 요인이 되는 것이다.

4 복수문서 요약의 특성

어떤 주제에 대하여 한 문서를 보는데 만족하지 못하고 여러 가지 문서를 찾아보고자 하는 이용자의 목적은 여러 문서가 특정 주제에 대하여 공통적으로 다루는 것과 다양한 견해로 차이를 두어 다루고 있는 것을 모두 파악하고자 하는 것이다. 그러므로 복수문서의 요약은 여러 문서 간에 공통점과 차이점을 모두 포함하고 있어야 이용자의 요구를 충족시켜 줄 수 있다.

■ 고유성(uniqueness)

복수문서 요약에서도 공통점을 파악하기 위해서는 단일문서 요약과 마찬가지로 중복성을 이용한다. 그러나 복수문서 요약이 가지는 특성은 중복성을 파악하는 것이 아니라 문서가 가지고 있는 정보의 고유성 또는 상이성을 파악한다는 점에 있다. 즉, 한 주제에 대하여 여러 문서가 얼마나 다양하게 다루고 있는지 비교하여 간략하게

핵심만 제공하는 것이 복수문서 요약의 주요 기능이 된다. 단일문서 요약에서 문서 내 한번만 나타난 정보는 큰 의미를 가지고 있지 못할 수 있으나 복수문서 요약에서는 특정문서에서 한번만 나타난 정보가 이용자에게 오히려 더 찾기 힘든 중요한 정보가 될 수 있다.

■ 정보량

복수문서는 단일문서보다는 요약대상 정보량이 매우 크다. 그러므로 10-30%의 압축률을 적용하는 단일문서보다는 압축률이 훨씬 낮아야 한다. 예를 들면 단일문서의 보편적인 압축률 10%를 5000개의 문장으로 구성된 문서집합 요약의 압축률로 적용한다면 요약의 제기능을 상실하게 된다. 또한 문서집합의 크기도 단일문서보다는 다양할 수 있다. 일반적인 주제를 다룬 복수문서의 요약을 보고 싶다고 하면 요약의 대상이 되는 문서집합이 매우 커지는 것이고 특수한 주제를 다룬 복수문서의 요약이라면 대상 집합이 상대적으로 매우 작을 수 있다. 그러므로 획일적으로 압축률을 적용하는 것보다는 섬세한 접근방식이 필요하다.

■ 개별문서간 이질성 처리

복수문서 요약의 특성은 문서집합내의 개별문서들이 가지는 이질적 요소들을 어떻게 처리할 것인가에 달려있다.

첫째, 문서의 길이 측면으로 보자면, 복

수문서 요약은 단일문서와는 달리 문서 하나의 길이는 크게 문제가 되지 않는 반면, 대신 문서집합내 각각 길이가 다른 문서를 조율해야 하는 복합적인 접근방식이 적용되는 특성이 있다. 즉, 길이가 아주 짧은 문서에서 추출한 정보와 아주 길이가 긴 문서에서 추출한 정보를 그대로 처리할 것인가 아니면 문서의 길이를 반영하여 처리할 것인가를 결정해야 한다. 예를 들어 문헌 내 용어빈도를 이용한 가중치를 사용하면 상대적으로 긴 문서가 유리할 수도 있기 때문이다.

둘째, 문서의 다양한 구조를 수용할 수 있어야 한다. 복수문서 요약의 대상은 같은 주제를 다루고 있지만 구조가 매우 다른 문서들이 될 수 있다. 이 경우 각 문서의 구조적 특성을 단일문서 요약시스템처럼 활용할 수 없을 뿐만 아니라 다양한 문서 구조를 교차 비교하여 평가할 수 있는 방안이 마련되어야 한다.

셋째, 다양한 언어가 복수문서 요약이 고려할 문서집합내 이질성이 될 수 있다. 단일문서가 여러 언어로 기술된 가능성보다는 문서집합내에 여러 언어로 기술된 문서가 출현할 가능성이 높다.

5 문서 환경에 따른 기능 비교

5.1 단일문서 요약의 주요 기능

■ 검색결과에서 개별 문서내용 파악

검색결과에서 적합한 문서를 골라내는데

가장 많이 사용된다. 이용자가 원하는 문서인지 알아보기 위해 색인어나 주제어를 사용할 수도 있지만 용어만 제공되면 문서 내에서의 용어간 관계가 드러나지 않는다. 그러나 요약에서는 개념의 관계가 분명하게 드러나기 때문에 원문의 정보를 파악하는데 효과적이므로 적합문헌 선정에 요약 을 활용하는 것이다.

■ 문서 클러스터링을 위한 기본 자질

문서를 주제별로 클러스터링하게 될 경우 원문서 전체를 대상으로 하면 시간적, 경제적 자원이 많이 소비된다. 만약 요약이 원문의 핵심 주제를 빠지지 않고 효과적으로 표현한다면 요약만을 가지고 클러스터링을 해도 유사한 결과가 나오게 된다.

■ 주제색인어 추출을 위한 대상

문서의 주제를 나타내는 주제어를 추출할 때 문서전체를 대상으로 하게 되면 상대적으로 주제연관성이 적은 용어도 추출될 확률이 높다. 이 경우 문서를 자동요약한 후 요약문내에서 색인어를 추출하게 되면 비주제어가 섞이게 될 가능성이 줄어든다(Kiyota, 2001).

■ 복수문서 요약의 선행과정

복수문서의 요약을 생성할 때 문서 전체를 사용할 수도 있지만 기존에 생성된 단일문서의 요약을 합쳐서 공통점과 차이점

을 비교, 추출하는 사례도 있다. 이 경우 문서비교에 드는 경제적, 시간적 자원을 최소화 할 수 있다.

5.2 복수문서 요약의 주요 기능

■ 문서집합의 주제요약

분류된 문서집합이나 검색결과에서 나타난 문서들의 주제를 요약하는데 활용된다. 웹 검색분야에서 검색결과로 많은 문서가 제공되었을 때 이용자들이 검색결과 내용을 신속하게 파악하는 것을 지원한다.

■ 정보갱신이 빈번한 분야의 동적요약 생성

통신사 뉴스처럼 같은 주제에 대해 연속적으로 새로운 정보가 발생할 때 정보발생 시점부터 가장 최근에 발생한 정보까지 추적하고 취합하여 하나의 요약으로 생성하는 경우이다. 이 경우는 최초 문서가 발생했을 때 기본 요약을 만든 후 추후 새로운 정보를 담은 문서가 입력될 때마다 비교하여 변경된 부분을 보완하고 새로운 정보를 추가하는 것이 일반적이다. 예를 들면 통신사 같은 경우 폭발사고를 난 날에는 거의 유사한 내용을 다루지만 조금씩 정보가 갱신된 뉴스문서가 수 십건 들어오는데 그 문서를 종합하여 보기는 쉽지 않은 일이다. 이때 한 사건을 기준으로 내용을 정리하는 복수문서 요약기능은 효과적인 방안이 될 수 있다(McKeown, 1999). 따라서 동적요

〈표 2〉 단일문서 및 복수문서 주요기능 비교

기능		분류기준			
		문서중심	이용자중심	지시적	통보적
단일문서	문서내용파악	◎		◎	
	주제색인어추출	◎			◎
	클러스터링자질	◎			◎
	복수문서요약자질	◎			◎
복수문서	문서집합내용파악	◎		◎	
	동적요약생성	◎			◎
	새문서 생성		◎		◎
	질의응답		◎		◎

약을 생성하는 복수문서 요약기능은 요약 연구분야의 중요한 부분을 차지하고 있다.

■ 새로운 문서 생성

여러 문서에 부분 부분 흩어져 있는 특정주제에 대한 부분을 모아 한 정보로 합쳐서 새로운 문서를 생성하는 기능이다. 복수문서의 이러한 기능을 활용하여 특정 인물의 바이오그래피 자동생성에 적용한 사례가 있다(Mani, 2001).

■ 질의 응답 기능

이용자가 관심있는 주제를 중심으로 복수문서에서 연관된 정보를 요약하여 제공하였을 때, 이용자가 원하는 정보를 요약 내에서 충분히 얻어 원문서를 필요로 하지 않게 된다면 이용자가 가진 질의에 대하여 복수문서 요약이 답을 제공한 것이라 볼 수 있다. 특히 이용자가 단답식 정보에 대한 질의를 한 것이 아니라 일정량 이상의 설명이 필요한 개념에 대하여 질의를 한

것이라면 단답식 위주의 정보추출에 치우쳐 있는 질의응답시스템보다는 복수문서 요약기능을 기반으로 한 질의응답시스템이 훨씬 효과적이다.

5 결 론

단일문서와 복수문서의 주요 기능을 요약을 분류하는 기준 중 요약의 기능과 최근 자동요약에서 관심이 높아져가고 있는 이용자지향성이라는 두 가지 요소를 가지고 비교하였다(표 2 참조). 요약기능 중 비평적 기능은 현재 자동요약에서는 보편적으로 적용되지 않고 있으므로 제외시켰다.

단일문서와 복수문서 주요 기능의 공통점을 찾아보자면 두 환경에서 모두 원문을 지시하는 기능보다 원 정보의 내용을 제공하는 도구로서 사용되는 성향이 강하다는 것을 알 수 있다.

이용자 지향성이라는 측면을 보면 단일문서 요약은 대부분 문서중심인 경우가 많아 다양한 이용자의 정보요구를 충족시키

기에는 적합하지 않는 것으로 나타났다. 그러나 복수문서 요약은 기존 정보에서 새로운 정보를 생산하고 이용자 질의에 답을 제공하는 것과 같은 보다 고차원적인 정보 활용을 위해 기능이 확장되어 가고 있는 것을 알 수 있다. 특히, 질의응답이나 새로운 정보원 생성과 같은 기능 등은 이용자가 원하는 주제에 맞추어 여러 문서에서 적절한 정보를 추출하여 제공함으로써 이용자의 정보요구를 1차적으로 해결할 수 있는 효과적인 방안을 제시하고 있다.

복수문서 요약은 원정보원을 대체하는 요약의 본래 목적이 아닌, 역동적인 문서 환경에서 새로운 문서 생성목적에 적합한 정보생산도구로 발전할 가능성이 있다. 이러한 가능성 때문에 최근 복수문서 요약의 특성과 자동분류나 인공지능 등 타 정보검색 분야와 결합한 다양한 시도가 나타나고 있고, 향후 정보 자동생성 연구의 기반이 될 것으로 기대된다.

참 고 문 헌

- Mani, Inderjeet. 2001. *Automatic Summarization*. Amsterdam: John Benjamins Publishing Co.
- McKeown, K. and D. Radev. 1999. "Generating summaries of multiple news articles" In Edited by I. Mani and M. Maybury ed. 1999. *Advances in Automatic Text Summarization*. Cambridge : MIT Press
- Kiyota, Y. and S. Kurohashi. 2001. "Automatic summarization of Japanese sentences and its application to a WWW KWIC index" *Proceedings of Applications and the Internet*, 120 -127.