

---

# 지식기반 회의관리 시스템 아키텍처에 관한 연구

김창수\* · 정희경\*\* · 이수연\*\*\*

## A Study on Knowledge based Conference Management System Architecture

Chang-Su Kim\* · Hoe-Kyung Jung\*\* · Soo-Youn Lee\*\*\*

### 요 약

본 논문에서는 모든 분야의 지식에 대해 온톨로지(ontology)화 하기보다는 회의 분야로 특화하여 회의 지식을 온톨로지화 할 수 있는 지식기반 회의관리 시스템 아키텍처를 제시한다. 또한, 시뮬레이션 및 분석을 통해 객관적 평가 기준을 마련한 다음 화상 회의, 음성인식, 행동 인식 등의 각종 인식 시스템을 통해 변환되고 가공된 정보를 체계화 하고 지식화하여 분석할 수 있는 시스템으로의 발전 가능성에 대해서 제시한다.

### ABSTRACT

This thesis proposes standard of knowledge-base system architecture for managing conferences in order to make ontology in the part of a conference rather than all parts. Also, this thesis proposes possibility of developing into the system that systematize transformed and processed information through various recognition systems, video conference, speech recognition, motion recognition, and so on, make knowledge and analyze it after preparing standards of objective estimation through simulation and analysis.

### 키워드

온톨로지, 회의시스템, 지식기반, XML

## I. 서 론

지식은 기본적으로 인간의 두뇌에 존재하기 때문에 객관적인 가치를 측정 및 평가하기가 어렵고 이러한 지식들을 시스템화하기에는 많은 정보에 의해 복잡도가 증대되는 문제점을 안고 있다. 또한 다양한 정보원들에서 생산되는 다양한 포맷의 콘텐츠들에 대한 이질성을 극복하고 상호운용성을 갖도록 연구되고 있다[1].

그러므로 지식관리에서 가장 중요한 요소인 지식들을 효과적으로 관리하며 체계화시키고, 분류/저장하며, 다

양한 관련 특성과 정보들을 이용하고, 지식들을 어떻게 사람/프로세스/활동과 맵핑시키며, 효과적으로 관리/유통될 수 있도록 할 것인가에 관한 필요성이 대두 되었다 [2,3].

현재까지의 지식관리 시스템은 효과적으로 지식 프로세스를 지원하고, 다양한 비정형 지식들을 통합하고 연동하는 방향으로 연구 및 발전되어 왔다. 그러나 일반적인 지식들을 모두 지식화 하기에는 모호성에 따른 복잡도 증가로 인해 지식 표준화에 한계를 가지고 있다.

이러한 지식기반 시스템중의 한 분야로써 회의 시스템

---

\* 청운대학교 인터넷학과

\*\* 배재대학교 컴퓨터공학과(교신저자)

\*\*\* 광운대학교 컴퓨터공학과

이 있는데, 본 논문에서는 기존 멀티미디어적인 측면에서 회의 시스템을 구성하는 일반적인 접근 방식을 탈피하고 회의 내용과 방식에 초점을 맞추어 기존 회의 시스템들을 그대로 흡수 할 수 있는 회의 내용과 진행방식에 대한 아키텍처를 제안한다. 이를 위해 회의 내용과 정보들을 XML(eXtensible Markup Language)로 구조화 한 다음 구조화된 지식에 대해 온톨로지화 과정을 거쳐 지식화 한다. 또한 회의가 진행되면서 미리 정의해 놓은 회의 모델을 이용하여 적합한 단어나 구문 등에 대해 온톨로지 패턴 매칭을 통하여 지능적인 회의록 구성이 가능하도록 한다. 그리고 현재의 자동 요약의 기술적 한계로 인하여 정확한 요약 및 정리를 위해 관리자가 직접 요약 과정에 개입이 가능한 형태로 구성하여 실용적인 아키텍처를 제안한다. 그리고 각 회의별로 최적의 회의 온톨로지 모델을 구성하여 효율적인 회의가 진행 될 수 있도록 한다.

## II. 관련연구

본 장에서는 지식기반 회의관리 시스템에 관련된 내용에 대해 살펴본다. 지식관리 시스템은 지식기반 시스템을 통칭하는 개념으로 본 논문에서 제시하는 지식회의 시스템의 경우는 의사결정지원 시스템에 해당하는 개념이다. 본 논문에서는 지능적 의사결정지원 시스템을 구체화하여 지식회의관리 시스템으로 구성한다. 또한 이러한 지식회의의 관리 시스템을 구현하기 위해 온톨로지, BSC(Balance Score Card), 지식 마이닝(mining)의 개념을 도입하여 지능화 할 수 있는 방안을 제시한다.

### 2.1. 지식기반 회의 시스템

현재 회의 시스템 연구는 크게 두 가지 방향으로 연구되고 있다. 그 중 한 가지 방법은 멀티미디어에 기반한 현재의 일반적인 오프라인 형태의 회의를 온라인상에서 쉽게 사용가능하도록 하기 위한 것이고, 다른 하나는 분석 통계에 따른 정보를 기반으로 회의를 규정하고 그 정보를 활용하는 형태이다.

본 논문에서는 멀티미디어 방식과 분석 및 통계 방식을 통합한 지식기반 회의 시스템 방식으로, 분석 통계적인 관점에서 출발하여 통계적으로 제시된 데이터를 지능적으로 분석하고 데이터에 대한 표준을 제시한다. 또한, 실제 회의 중에도 실시간으로 정리 및 지식화가 가능하여

좀 더 효율적인 회의 진행이 가능하게 된다.

### 2.2. 온톨로지

온톨로지는 특정 응용영역에 대해 만들어지는데, 그 분야의 기본 개념에 대한 정의와 그들 간의 관계에 대한 명세로 이루어진다. 본 논문에서는 회의 온톨로지에 초점을 맞추어 XML, 온톨로지 논리 추론 기술, 분류 체계관리 기술, 지식 표현 기술, 색인 기술들을 위주로 하여 온톨로지를 구축한다. 또한, 회의 정보의 지식화에 초점을 맞추어서 논제를 지능적으로 온톨로지화 하며, 이렇게 온톨로지화된 정보를 XML로 구조화하고 모델화 하여 관리할 수 있는 방안을 제시하여 기존 의사결정의 장점을 포용함과 동시에 좀 더 시각적이면서도 효율적이고 구조적인 지식기반 회의관리 시스템을 설계 및 구현한다.

### 2.3. 기존 회의관리 시스템의 문제점

기존 멀티미디어 기반의 온라인 회의시스템의 경우는 시공간의 제약 없이 회의가 가능하다는 장점은 있지만 실제적인 회의에 대한 지식화가 불가능하다는 단점을 가지고 있어 회의 참석자들의 개별적인 역량에만 의존하여 회의가 진행되도록 하고 있다.

그리고 일반적인 지식관련 시스템들의 경우는 언어처리 및 지식표준화의 어려움으로 인하여 지식처리에 지능화 부분을 적용하는 연구는 미비한 실정으로 주로 지식베이스를 구축하려는 노력이 진행 중이다. 그러나 회의 지식을 지식화 하기 위해서는 전문가에 의해 지식 분류가 수작업으로 이루어져야 하므로 관점에 따른 오분류와 함께 상당한 관리 비용이 필요하다[4]. 그러므로 이런 부분을 지능적으로 수행할 수 있는 지능적 회의 시스템이 필요하게 되었으며 회의 지식에 대한 객관적인 평가와 측정에 대한 관심도 필요하다.

또한, 일반적인 회의 시스템의 경우는 회의가 진행되면서 공유를 위한 표현방법이 상당히 제한적이며 시간의 흐름에 따라 회의 내용을 나열식으로만 관리되어 나중에 회의내용에 대한 조회나 검색 시 일일이 해당 내용을 모두 검색해야 하는 문제점을 안고 있다[5].

## III. 지식기반 회의 시스템 아키텍처

그림 1은 기존의 오프라인 및 멀티미디어기반 회의시

시스템들의 문제점들을 보완하여 제안하는 지식기반 회의 관리 시스템의 전체 아키텍처이다. 전체적으로 회의 서버와 회의 클라이언트로 구성되며 서버와 클라이언트는 TCP/IP로 소켓통신을 통해 XML 메시지로 교환된다. 회의 서버에서는 XML 메시지를 받은 후 지식화 과정을 위해 전처리를 하며 로그를 기록한 다음 온톨로지 추출기를 통해 온톨로지로 추출이 가능한 것을 추출하고 처리하지 못한 내용에 대해서는 예외 처리기에서 처리한다.

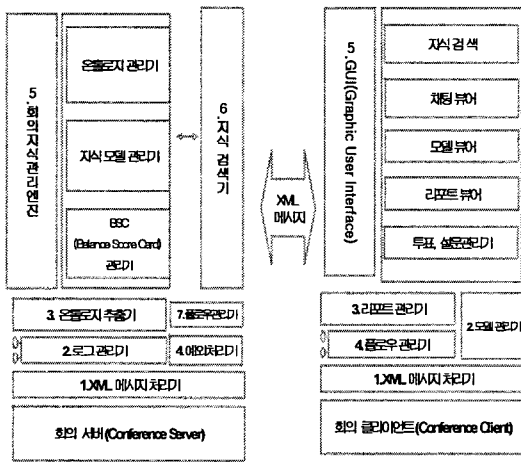


그림 1. 지식기반 회의 시스템 전체 아키텍처  
Fig. 1. The total architecture of ontology-based conference system

회의 클라이언트에서는 XML 메시지를 전송하게 되며 모델 관리를 통해 적합한 모델 선택 후 플로우 및 리포트 관리기와 연동하여 GUI를 통해 보인다.

### 3.1. 회의 서버 아키텍처

#### 3.1.1 XML 메시지 처리기

회의에 참석한 사람과 진행자는 지식기반 인터페이스를 가지는 클라이언트를 사용하여 온톨로지를 이용한 지식맵(map)을 작성할 수 있다. 작성된 지식맵 데이터는 데이터 부분만 회의 서버로 전송되며 XML 형태로 구조화되어 일반 회의정보, 선호도, 의사결정 순서정보, 지식기반 구조화 정보 별로 저장되고 관리된다.

회의 서버에서는 XML 메시지를 입력 받아 해당 메시지의 내용을 파싱하여 메시지와 회의 정보 및 내용을 분리해 낸다. 이러한 정보들을 클라이언트 요청에 의해 꾸준히 전송한다.

#### 3.1.2 로그 관리기

로그 관리기는 서버에서 입/출력되는 데이터에 대한 접속로그, 에러로그, 정보로그로 분리하여 기록된다.

접속로그의 경우 추후 접속 통계관련 자료로 사용되며, 에러로그는 에러 상황의 모니터링에 사용되고 정보로그의 경우는 회의 정보와 회의 내용 등을 기록하여 자동 지식화 엔진에서 처리하지 못한 내용 등을 담고 있어 나중에 사용자가 직접개입이 가능하도록 데이터를 로그형태로 기록한다.

#### 3.1.3 온톨로지 추출기

논문에서는 특정한 카테고리 분류되는 문서들에 공통적으로 자주 나타나는 문장 형태와 키워드들을 온톨로지 패턴으로 만들어 온톨로지 패턴 매칭에 의해 문서 분류하는 방법을 제시하고, 통계적인 문서 분류에 온톨로지 패턴들을 이용한 분류 방법을 접목시킴으로써 분류의 정확도를 높이는 방법에 대해 제안한다. 또한, 온톨로지 패턴 매칭의 경우는 빠른 처리속도를 보장함으로써 실시간 회의 시스템에 적합하다. 그림 2에 이를 나타낸다.

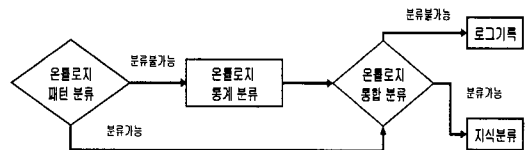


그림 2. 온톨로지 하이브리드 분류 방식  
Fig. 2. division method of ontology hybrid

논문에서는 분류하려는 구문과 분류 대상 범주들을 온톨로지 벡터로 구성하고, 두 온톨로지 벡터 사이의 유사한 정도를 비교하여 유사도가 가장 높은 온톨로지로 구문을 분류하는 벡터 유사도 이용 방법[4,6,7]을 온톨로지 통계적인 분류방법에 사용하였다. 온톨로지 벡터로 표현된 구문 P와 카테고리 Cj의 유사도는 수식(1)과 같이 계산된다.

$$\text{Similarity}(P, C_j) = \cos \theta = \frac{P \cdot C_j}{|P| |C_j|} \quad \text{수식(1)}$$

구문과 카테고리의 벡터 표현에서 각 색인어의 가중치로는 일반적으로 많이 사용되는 역문헌빈도[6,8]대신 카

태고리의 분리 능력이 우수한 색인어에 높은 가중치를 주는 역카테고리빈도[9]를 사용하였다.

### 3.1.4 예외 처리기

회의가 진행되면서 실제적으로 지능화 단계에서 처리가 불가능한 데이터 및 메시지에 대한 유효화와 프로세스 및 회의 온톨로지간의 추론 결과에 대한 정의가 필요하다.

또한, 회의가 진행되면서 발생하는 각종 프로세스적인 부분의 예외 처리모형이 필요하다. 따라서 회의 프로세스에서 예외 상황이 발생할 경우 이를 공지하는 역할과 공지를 받아서 보고하고 관리하는 역할로 구분하여 예외처리가 진행된다. 이렇게 예외사항이 발생할 경우 기본적으로 로그기록이 처리되며 이 중에서 관리자(사회자)가 실시간/비 실시간으로 내용을 요약 정리할 수 있는 방안을 마련하였다.

### 3.1.5 지식 검색기

회의 지식 관리 엔진을 통해 온톨로지화된 지식들에 대해서 검색하는 것으로 한정하며, 현재 회의 지식 관리 엔진을 통해 지식화 과정을 거친 데이터의 경우 이미 지식화 되어 있어 빠른 검색속도를 보장한다.

### 3.1.6 플로우 관리기

플로우 관리기는 회의 프로세스 개선/회의 프로세스 정의, 해석 및 실행/회의 정의 및 실행/회의 태스크(task) 목록관리/회의 감시, 추적, 감사/회의 경고 및 알림 등을 관리한다. 표 1은 회의 플로우 관리 XML 문서의 예를 나타낸다.

## 3.2. 회의 클라이언트 아키텍처

클라이언트의 XML 메시지 처리기의 경우 서버의 XML 메시지 처리기에 해당 회의 내용과 회의 정보를 XML 구조로 전송하는 역할을 담당한다.

모델 관리기는 서버 쪽에서 넘어온 모델 정보에 뷰잉(viewing) 정보를 포함하여 실제 GUI 적으로 표현하기 위한 정보들을 관리한다.

리포트 관리기는 지식화된 자료들을 텍스트 및 그래픽 등 다양한 형태로 리포팅이 가능하도록 관리한다.

플로우 관리기는 서버쪽 플로우 메시지를 받아 전체 회의에 대한 진행 순서, 알림, 경고 메시지 등에 대한 처리를 담당한다.

표 1. 회의 플로우 관리 XML  
Table 1. conference flow management XML

```
<info>
  <seq>00000001</seq>
  <!--type: seq, doncare, repeat, any, every -->
  <type>seq</type>
  <due_date>20050503</due_date>
  <due_time></due_time>
  <!--status : start, close, abnormal-close, suspend, reserve -->
  <status>start</status>
  <flow_name><![CDATA[회의 제목]]></flow_name>
  <session seq='1'>
    <date>20050504</date>
    <time>1420</time>
    ...
  </session>
  <session seq='2'>
    ...
  </session>
```

## 3.3. 협업 아키텍처

기업에서 구성원들 간의 원활한 의사소통이 가능하게 하기 위해 많은 노력들이 있었으나, 현재의 방법들은 회의 프로세스에 대한 확인 및 검증이 어려울뿐더러 표준의 변화에 따른 시스템의 변경과 예외 처리 등의 변화가 응용성을 저해시키는 요인이 되고 있다. 이를 해결하기 위해서는 예외 처리를 함께 고려하여 메타모형 및 에이전트를 기반으로 협업이 가능한 아키텍처로 구성하는 것이 우선되어야 한다.

그림 3은 일반적인 회의 형태를 개선하고자 제안하는 시스템의 협업 아키텍처를 표현한 것이다. 일반적인 회의 시스템에서 회의 참석자 역할은 동일하지만 ICoMS(Intelligent Conference Management System)에서 서기의 역할과 함께 의장의 일부 역할도 함께 포함한다. 회의 참석자들은 자유롭게 의견을 개진하고 투표에 참여하며, 모든 과정에서 ICoMS를 통해 실시간으로 분석/정리/공유된다. 이때 의장 및 서기의 정리 및 표기에 따른 부담은 사라지며 ICoMS에서 자동으로 처리되지 않은 것들에 대해서만 직접 처리하는 방식이다.

이렇게 될 경우 회의가 진행되는 동안 의견 개진 → 정리 → 결정 → 분석 → 공유 → 리포팅이 실시간으로 신속하게 처리가 되므로 효율적이고 빠른 협업이 이루어질 수 있는 것이다.

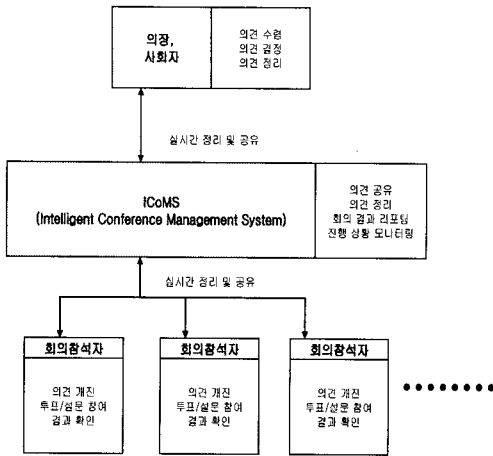


그림 3. 제안 시스템의 협업 아키텍처  
Fig. 3. The cooperation architecture of proposal system

#### IV. 시뮬레이션

##### 4.1. 시스템 전체 구성

지식기반 회의관리 시스템에서 서버환경은 펜티엄 III 1GHz, 윈도우 2000 서버 환경에서 자바를 이용하여 개발되었다. 클라이언트 환경은 펜티엄 III 1GHz, 윈도우 2000 워크스테이션에서 Visual C++를 이용한 ActiveX Control로 시뮬레이션 되었다. 그림 4는 지식기반 회의관리 시스템의 전체 시스템 흐름도를 나타내고 있다.

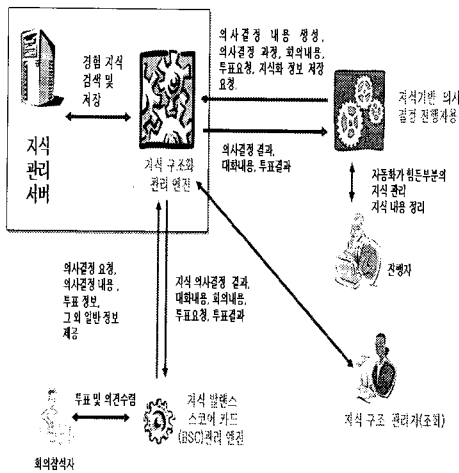


그림 4. 전체 시스템 흐름도  
Fig. 4. The total system flowchart

채팅 내용은 채팅 뷰어를 통해 보일 수 있도록 구성하며, 현재 메시지를 입력한 사용자의 메시지가 중앙에 큰 창으로 하여 집중도를 높이기도 하였다.

모델 뷰어는 본 시스템이 제안하는 중요한 UI(User Interface)중 하나로써, 회의별로 다양한 모델 형태가 있을 수 있으며, 이를 그래픽 적으로 표현해 주어 표현의 한계가 없도록 하였다.

그림 5는 모델에 대한 뷰 형태를 나타내고 있으며 그래픽적으로 표현가능 하므로 다양한 형태로의 변형도 가능하다.

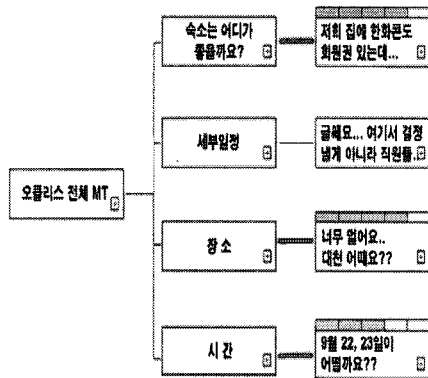


그림 5. 모델 뷰어  
Fig. 5. The model viewer

그림 6은 리포트 뷰어 기능으로서 기본적으로 인쇄기능을 포함하고 있으며 회의록 형식을 규정하여 다양한 형태의 보고서가 가능하도록 구성한다.

회의록				
OOO 전략기획팀			OOO 기업	
부장	과장	대리	PM	DAO
회의시간	2004년 4월 20일 오전 10:00 ~ 11:00		회의장소	
참석자	유XX 부장, 김XX 과장, 김XX 대리			
회의제목				기록작성자
회의내용				

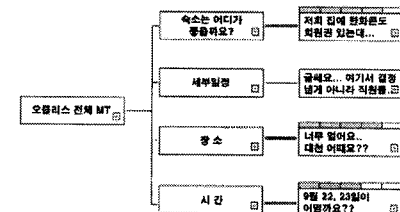


그림 6. 리포트 뷰어  
Fig. 6. The report viewer.

그림 7은 투표 및 설문 결과에 대해 그래프 형태로 위단에 표현하고, 아래에는 간략한 텍스트를 두어 해당 내용에 대한 신뢰도 및 결정 내역에 대해 일목요연하게 표현하도록 구성한다.

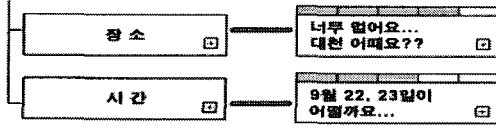


그림 7. 투표, 설문내용 및 결과 보기  
Fig. 7. Representation of voting, research content, result

### 4.2 회의 서버 시물레이션

회의 서버에서는 XML 메시지 처리기를 통해 모든 정보가 회의 클라이언트와 통신하며, 들어온 메시지는 기본적으로 로그 관리기에서 로그 기록으로 남기게 된다. 그런 다음, 지식 추출기를 통해 자동으로 요약정리가 가능한 지식에 대한 추출이 이루어진 다음, 회의 지식 엔진을 통해 종합적으로 체계화되고 관리된다. 이때 자동으로 정리가 안 된 부분은 예외처리를 통해서 별도 처리를 하도록 하며 지식 검색기를 통해 지식을 검색할 수 있다. 또한, 별도의 플로우 관리기를 두어 회의진행시에도 관리될 수 있도록 하였다.

### 4.3 회의 클라이언트 시물레이션

회의 클라이언트는 실시간 온라인 회의에서 회의 내용을 시각적으로 표현하는 시스템으로, 회의 참석 인원이 대화 창에서 타이핑하는 내용을 회의 진행자가 대화창 옆에 붙어있는 다른 화면에 회의 주제를 중심으로 표현해 준다. 사용자들이 일일이 입력하는 모든 내용을 그대로 요약 한다는 것은 상당히 비효율적인 작업이다. 그리고 패턴 매칭을 통해 지능적인 요약이 이루어지고 난 다음, 패턴 매칭이 이루어지지 않은 내용에 대해 회의 진행자가 직접 중요한 요점에 해당하는 부분을 드래그앤드롭(drag & drop) 방식으로 쉽게 간편하게 요약 정리할 수 있다.

그림 8은 진행된 회의내용과 이미 선택된 주제에 대한 화면으로 요약해서 보여준 화면이다. 이렇게 정리된 각 주제별로 오른쪽 하단의 '+' 버튼을 클릭하면 선택되지 않은 다른 내용들이 보여지며, 어떠한 내용 중에 해당 주제가 선택이 되었는지 알 수 있다.

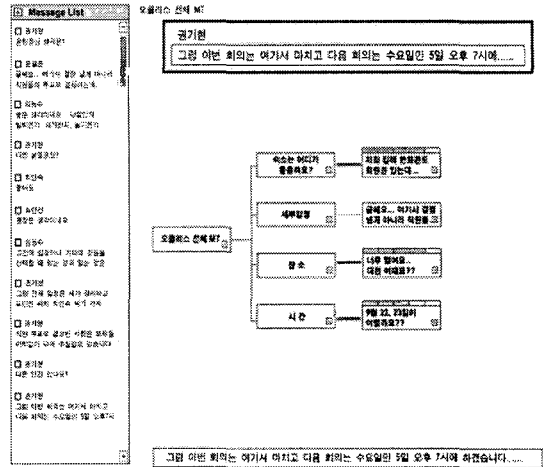


그림 8. 회의 내용 요약  
Fig. 8. The abstract of conference content

선택된 주제에 대해 결정된 지식에 대해서만 보여질 수도 있으며 그 이외의 의견에 대해서도 일목요연하게 볼 수가 있다. 이렇게 보여주는 이유는 결정된 의견일 경우 신뢰도가 가장 높으며 가장 관심이 되는 중요한 데이터이고 그 외의 데이터는 해당 지식의 결정에 대한 참조 데이터이기 때문이다.

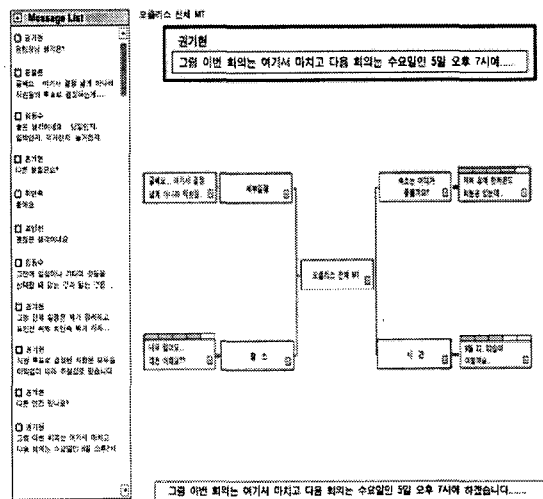


그림 9. 다양한 표현 및 보고 기능  
Fig. 9. The multiplicity representation and Reporting function

그림 9는 그림 8에서와 같이 요약된 형태로 표현이 가능한 회의 내용의 경우 간단한 선택만으로 위와 같은 보고서의 형태가 바뀔 수 있다. 회의 내용에 적절한 형태를 미리 제공을 해줌으로써 회의 내용을 집중력 있게 진행할 수 있으며, 이렇게 생성된 데이터는 자동적으로 요약 정리되고 다양한 형태로 표현이 가능하다.

#### 4.4. 평가 분석 및 고찰

##### 4.4.1 평가 분석

본 논문에서는 시물레이션을 위한 문서 집단으로써 한겨레신문(<http://www.hani.co.kr>) 웹사이트로부터 수집한 100개의 뉴스 기사를 이용했으며, 뉴스의 경우 일반적인 다른 문서에 비해 정형화되어 있으며 명확한 표현들을 사용하고 있어서 실험 데이터로 사용하였다.

뉴스를 분석하여 대분류 6개와 소분류 30개를 선정하였으며, 선정 방식은 경험적으로 선택하였다. 시물레이션에 사용한 구문 패턴은 총 80개였으며, 키워드 집합은 한 단어로 구성된 집합이 78개, 두 단어로 구성된 집합이 92개, 세 단어 이상으로 구성된 집합이 80개였다.

그리고, 첫 번째 시물레이션에서는 패턴을 이용할 구문패턴 방식과 키워드 집합방식을 적용하여 분류율과 정확도를 구했으며, 특히 키워드 집합을 이용한 방법에 대해서는 집합 내 단어 수에 따른 정확도 변화를 구하였다. 두 번째 시물레이션에서는 같은 문서집단을 대상으로 통계적 방법을 이용한 분류의 정확도를 구하고, 복합적인 방법에 의한 분류 실험하여 정확도를 비교하였다. 본 시물레이션에서 측정된 분류율과 정확도의 정의는 패턴에 의한 분류의 정확도를  $P_p$ , 분류율을  $C_p$ , 통계적 방법에 의한 분류의 정확도를  $P_s$ , 분류율을  $C_s$  라고 할 때 수식(2)와 같다.

$$\begin{aligned} \text{정확도}(PP, PS) &= \frac{\text{정확하게분류된문자수}}{\text{분류된총문서수}} \\ \text{분류율}(CP, CS) &= \frac{\text{분류된총문서수}}{\text{총문서수}} \end{aligned} \quad \text{수식(2)}$$

위와 같이, 정확도는 분류된 문서 중에서 정확하게 분류된 문서의 비율을 의미한다. 하지만 정확도만으로는 분류 자체의 오분류에 대해서는 고려하기 어렵다. 분류율과 정확도는 서로 반비례 관계가 있으므로 이를 정확히 표현하기 위해서는 분류율과 정확도만으로는 충분치 않다. 그래서 Lewis 등은 분류율과 정확도의 문제점을 보완하기

위해 F-measure 척도를 제안하였다[7]. 수식(3)은 F-measure 척도에 대한 정의식이다.

$$F\beta = \frac{(\beta^2 + 1) \cdot \text{정확도} \cdot \text{분류율}}{\beta^2 \cdot \text{정확도} + \text{분류율}} \quad \text{수식(3)}$$

F-measure에서  $\beta$ 는 분류율과 정확도의 중요도를 반영하기 위한 가중치를 의미한다. 즉,  $\beta = 0$ 이면 F-measure 값은 정확도와 동일하고,  $\beta = \infty$ 이면 F-measure 값은 분류율과 동일하다.  $\beta = 1$ 이면 분류율과 정확도에 동일한 가중치를 적용하여 F-measure 값을 계산한다.

따라서 F-measure를 분류율과 정확도의 중요도에 따라  $\beta$  가중치를 선택적으로 조정할 수 있으며 본 논문에서는  $\beta = 0.1$ 로 하여 회의라는 성격상 분류율 보다는 정확도에 더 높은 가중치를 두어 적용하였다. 이로써 분류율과 정확도를 회의 시스템에 적합하도록 함께 분석이 가능하여 정확도 또는 분류율만을 가지고 분석하는 것 보다는 객관적인 시스템의 성능 측정이 가능하다.

##### 4.4.2 실험 결과

표 2. 패턴에 의한 분류 결과  
Table 2. The division result by pattern

분류방법	표본 집단		
	분류율	대분류	
		정확도	F-measure
구문패턴	36.7%	89.4%	88.15%
키워드집합 3개	35.8%	93.3%	92.35%
키워드집합 2개	44.1%	92.2%	91.21%
키워드집합 1개	64.1%	83.2%	82.96%

표 2는 표본 추출된 뉴스를 대상으로 한 분류 시물레이션 결과이며 구문패턴에 의한 방식과 1,2,3개의 키워드 집합에 따른 분류율, 정확도, F-measure를 나타내고 있다.

결과를 보면 패턴을 이용한 분류는 두 가지 모두 분류율이 대략 30~64%인 것으로 보아 낮았으나 정확도는 83%이상으로 대부분 매우 높은 것으로 나타났으며 F-measure 값도 82%이상으로 나타났다.

또한, 키워드 집합을 사용한 경우에는 키워드 집합 내 매칭 되는 단어의 수가 많아질수록 분류율은 낮아지지만 정확도와 F-measure 값은 높아짐을 알 수 있다.

이렇듯, 키워드 집합을 이용한 패턴인식이 구문패턴에 의한 방식보다는 4~6%정도 높은 정확도와 F-measure값을 나타내고 있으나 구문패턴의 경우 초기 도메인 온톨로지의 구축 정도에 따라 그 정확도는 많이 향상될 수 있는 부분이 있다.

표 3. 통계적인 방법에 의한 분류 결과  
Table. 3 The division result by statistical method

분류방법	표본 집단		
	분류율	대분류	
		정확도	F-measure
통계적방식	42.9%	80.6%	79.90%
구문패턴+통계적방식	32.7%	91.9%	90.28%
키워드집합+통계적방식	30.6%	97.3%	95.24%

표 3은 통계적인 방법에 의한 분류(Stat)의 정확도와 통계적인 방법에 본 논문에서 제시한 패턴에 의한 분류 방법을 접목시킨 복합적인 방법의 정확도와 F-measure 값을 나타낸다. 통계적 방식의 경우 단어빈도에 따라 분류되도록 하였으며, 키워드 집합의 경우에는 가장 정확도가 높았던 키워드를 3개 사용한 경우를 적용하였다. 이 결과를 보면 우선 통계적인 방법의 분류율은 패턴에 의한 방식보다는 높았으며, 본 논문에서 제안하는 하이브리드 방법을 사용한 경우에는 패턴에 대한 분류를 적용한 다음의 통계적 방식의 분류를 적용하는 방식이어서 분류율 면에서는 오히려 다소 떨어지는 면을 보였다. 그러나 정확도와 F-measure 값은 향상되는 것을 알 수 있다.

일단, 통계적인 분류 방법만을 사용할 경우는 패턴만을 이용한 결과와 비교하여 정확도와 F-measure값은 약 9~11% 정도 낮은 것을 알 수 있으며, 본 논문에서 제안한 복합적인 방법을 사용했을 경우, 구문 패턴을 이용한 방법은 2.5~3.5% 정도 정확도와 F-measure값이 향상되었고, 키워드 집합을 이용한 방법은 4~5% 정도 정확도와 F-measure값이 향상되었다.

### V. 고찰 및 결론

본 논문에서는 지식기반 회의 관리 시스템 아키텍처

를 제안하면서, 통계적인 문서 분류에 문장의 의미를 파악할 수 있는 패턴들을 이용한 분류 방법을 접목시킴으로써 분류의 정확도를 높여서 회의 진행자가 수작업으로 분류하고 정리하는 부분들을 최대한 자동화하기 위한 방법을 제안하였다.

패턴의 표현 및 적용 방법으로는 패턴 매칭에 사용할 문장 형태를 구문 패턴에 의한 방법과 패턴 매칭에 사용할 문장 형태를 단어들의 집합으로 나타낸 키워드 집합에 의한 방법을 제안하였으며, 구문 패턴에 의한 방법과 키워드 집합에 의한 방법을 각각 통계적인 문서분류 방법과 접목시켜, 신문에서 표본 추출된 뉴스를 대상으로 문서의 자동 분류 시뮬레이션을 수행하였다.

그 결과, 문장의 의미를 파악할 수 있는 패턴에 의한 분류 기법을 통계적 기법에 접목시킨 경우가 통계적인 분류 방법만을 사용한 경우 보다 높은 정확도를 나타내었다.

또한 멀티미디어적인 방식에 대해서는 일반적인 하나의 인터페이스로만 국한하고 회의 지식 부분에 초점을 맞추어 회의의 진행 자체에 대한 지식화 방안에 대해 구체적으로 제시하였다. 그에 대한 내용으로 회의 내용을 지식화 하여 회의 온톨로지에 대한 구축 방법에 대해 제시하였으며 각각의 회의에 대해 모델 개념을 도입하여 회의의 성격에 적합한 모델을 선택하여 회의가 진행될 수 있도록 하였다.

본 시스템에서는 온톨로지 지식모델을 통하여 좀 더 효율적인 지식기반 회의 관리 시스템을 구현해 보았으며, 논제에 따른 일반 정보와 지식의 통합 그리고 관점의 표현을 온톨로지의 표현방법인 개념과 관계로서 구현하였다.

그리고 지식화된 온톨로지에 대한 객관적인 평가나 측정에 대한 부분은 BSC의 도입으로 지식을 정량화하여 객관적인 지표로 활용이 가능하도록 하였다.

또한, 마인드 맵 형태의 지식맵을 온톨로지 맵을 회의에 참석한 모든 사용자들이 공유할 수 있도록 하여 빠른 공유가 가능하며 추후에 회의 내용을 검색 할 때도 온톨로지화가 실시간으로 진행되어 자기가 원하는 내용에 대해서 빠른 조회가 가능하다.

회의가 진행되면서 자동으로 요약이 가능한 부분은 온톨로지 추출기를 통해 지능화하여 자동적으로 회의지식을 온톨로지화 하여 의장(관리자)의 역할을 최소화 했으며 회의의 성격상 정확도가 많이 요구되므로 직접 사용자가 온톨로지화 과정에 참여하여 온톨로지를 컨트롤 가능



하도록 인터페이스를 지원하고 회의가 종료되고 난 다음 바로 다양한 리포팅이 가능하도록 했다.

향후 과제로는 현재의 텍스트기반의 데이터가 아닌 멀티미디어 데이터에 대한 온톨로지화도 연구 가능성이 충분하며 두 개 이상의 패턴이 매칭되어 모호성이 발생한 경우에 어떤 카테고리로 분류할 것인가를 결정하는 방법, 수작업으로 해왔던 패턴을 찾아내는 작업을 자동화하는 방법 등에 대한 연구가 필요하다.

### 감사의 글

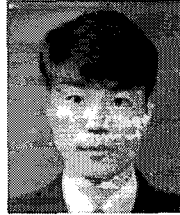
본 연구는 2005년도 광운대학교 교내연구비의 지원에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

### 참고문헌

- [1] J. Davies et al, "On-To-Knowledge: Semantic Web Enabled KnowledgeManagement", In Web Intelligence. Ning Zhong (ed), Springer, pp. 277-300, 2003
- [2] "Innovative Approaches for Improving Information Supply", Gartner Group Report, September 2001, M-14-3517
- [3] Robert Perry, and Robert Lancaster, "Enterprise Content Management: Expected Evolution or Vendor Positioning?", Yankee Group Report, June 2002
- [4] 문현정, "XML 구조적 특징을 이용한 온톨로지 기반의 지식 탐사 모델", 창원대 대학원, 2003
- [5] Daniel K.Power, "Decision Support Systems: Frequently Asked Questions", DSSResources.com, 2004
- [6] N. Kushmerick, E. Johnston, and S. McGuinness, "Information Extraction by Text Classification", IJCAI-01 Workshop on Adaptive Text Extraction and Mining", 2001
- [7] T. Lau and Y. Sure, "Introducing Ontology-based Skills Management at a Large Insurance Company", Proc. Modellierung, Germany, March 2002
- [8] Jerry R.Hobbs, Douglas Appelt, John Bear, David Israel, and Mabry Tyson "FASTUS: A System for Extracting Information from Natural-Language Text".
- [9] 강승식, "한국어 형태소 분석기와 한국어 분석 모듈 (HAM: HangulAnalysis Module)", 국민대학교, 2003

### 저자소개

#### 김 창 수(Chang-Su Kim)



1996년 배재대학교 전자계산학과 (이학사)  
1998년 배재대학교 전자계산학과 (이학석사)

2002년 배재대학교 컴퓨터공학과(공학박사)  
2001년~2004년 배재대학교 IT 교육센터 책임강사  
2005년~현재 청운대학교 인터넷학과 전임강사  
※ 관심분야 : XML, ebXML, Semantic web, 멀티미디어 문서정보처리, u-Logistics

#### 정 희 경(Hoe-Kyung Jung)



1985년 광운대학교 컴퓨터공학과 (공학사)  
1987년 광운대학교 컴퓨터공학과 (공학석사)

1993년 광운대학교 컴퓨터공학과(공학박사)  
1994년~현재 배재대학교 컴퓨터공학과 교수  
※ 관심분야 : 멀티미디어 문서정보처리, XML, SVG, Web Services, Semantic Web, MPEG-21유비쿼터스 센서 네트워크

#### 이 수 연(Soo-Youn Lee)



1969년 광운대학교 통신공학과(공학사)  
1979년 연세대학교 전자공학과(공학석사)

1983년 교토대학 정보공학과(공학박사)  
현재 광운대학교 컴퓨터공학과 교수  
※ 관심분야 : SGML, XML, 데이터베이스