

군 정보통합을 위한 메타데이터 기반의 데이터 그리드 시스템

나민영*

A Metadata-based Data Grid System for the Integration of Military Information

Minyoung Ra*

요약

최근 들어 유비쿼터스 시대를 맞이하여 군 관련 정보의 통합운용이 점점 더 중요한 주제로 부각되고 있다. 기존 정보시스템을 통합 운용함에 있어 문제의 핵심은 정보의 시맨틱의 불일치이다. 이의 해결을 위해 메타데이터의 활용이 요구되고 있다. 메타데이터는 데이터에 관한 데이터를 말하는 것으로 자원의 검색을 빠르게 하는데 사용되며 또한 시맨틱 갭을 해결하는데 사용된다. 본 논문에서는 군 정보통합을 위한 메타데이터 기반의 데이터 그리드 시스템을 다룬다. 먼저 군 정보 통합을 위한 접근방법으로서 연합 데이터베이스와 데이터 그리드에 관하여 살펴보고, 메타데이터 기반의 데이터 그리드를 다룬 다음, 이에 기초한 군사자료 데이터 그리드 시스템을 제시한다.

Abstract

Recently, the need for integration of military-related information is increasing. The fundamental problem for integrating existing information systems is semantic inconsistency. To solve this problem, the use of metadata is essential, because metadata is a key component for effective information integration. In this paper, we present a metadata-based data grid system for integrating military data sources. First, we investigate federated database approach and data grid approach for information integration, and deal with metadata-based data grid. Then we present a data grid system for the integration of military information.

▶ Keyword : 메타데이터(Metadata), 정보통합(Information Integration), 데이터 그리드(Data Grid)

• 제1저자 : 나민영

• 접수일 : 2008. 1. 21, 심사일 : 2008. 2. 13, 심사완료일 : 2008. 2. 17.

* 육군사관학교 전자정보학과 교수

I. 서론

대용량 데이터에 관한 요구가 많아짐에 따라 정보통합에 대한 관심이 계속 증가하고 있다.[1,10,11] 정보통합은 다수의 자치적이고 이질적인 데이터 소스에 걸쳐 질의 수행이 요구되는 응용 분야에서 당면한 큰 도전으로 여러 데이터 소스를 가진 조직에서는 필수적으로 요구된다. 특히 최근 들어 유비쿼터스 시대를 맞이하여 정보의 상호운용이 점점 더 부각되는 이 시대에 정보의 통합은 매우 중요하다. 군도 예외는 아니어서 날로 늘어나는 대규모 정보의 통합 운용에 대한 관심이 날로 증가하고 있는 실정이다. 특히 각 세대별로, 또는 각 군별로 구축된 각종 데이터 소스들은 매우 이질적이어서 이들의 통합은 매우 절실하다고 할 수 있다.

데이터베이스 시스템 엔진 관점에서의 정보통합은 메타데이터 스키마에 대한 질의처리 문제로 볼 수 있는 반면, 사용자 관점에서 보면 정보통합은 질의처리를 가능하게 하는 매핑에 좀 더 초점을 맞춘다고 할 수 있다. 기존 시스템의 통합 운용에 있어 문제의 핵심은 정보의 시맨틱의 불일치이다. 이의 해결을 위해 메타데이터가 요구되고 있다. 메타데이터는 데이터에 관한 데이터를 말하는 것으로 자원의 검색을 빠르게 하는데 사용되며 시맨틱 갭을 해결하는데 사용된다. 메타데이터 관리란 어떤 정보를 가지고 있고, 그것이 어디에 있으며 또 어떤 가치를 갖고 있는가를 사용자에게 제공하게 된다.

데이터 그리드는 정보통합을 이루기 위한 새로운 접근방법으로서 네트워크를 통하여 대규모 데이터의 공유와 접근을 가능케 해준다.

본 논문의 구성은 다음과 같다. 2장에서는 군 정보통합을 위한 접근방법으로서 연합 데이터베이스와 데이터 그리드에 관하여 살펴보고, 3장에서는 데이터 그리드 개념과 메타데이터 기반의 데이터 그리드를 다룬다. 4장에서는 메타데이터 기반한 군사자료 데이터 그리드 시스템을 제시하고, 5장에서 결론을 맺는다.

II. 정보통합을 위한 접근방법

이 장에서는 먼저 정보통합의 수준에 관하여 살펴보고 정보통합을 위한 접근방법인 데이터 연합에 의한 방법과 데이터 그리드에 의한 방법을 살펴본다.

2.1 정보통합의 수준

정보통합의 수준은 투명에 의해 구분된다. 즉, 투명이라는 정도(종류)까지 지원되는나이다. 투명(transparency)은 정보통합에 있어서 가장 핵심적인 가치로 정보통합 시스템 평가에 대한 주요 척도이기도 하다. 다음은 정보 통합시 제공될 수 있는 투명의 종류들이다.[15,16] 대부분의 시스템은 이중 일부만을 지원하므로 시스템에 따라 그 목적에 맞는 투명을 지원하게 된다.

- 이질(Heterogeneity) 투명 : 데이터 소스를 액세스하는 애플리케이션은 그 소스의 실제 구현(예를 들어, 실제 구현이 파일시스템이든, DB2이든, 혹은 오라클 DBMS 이든 간에)과는 무관해야 하는 능력
- 이름(Name) 투명 : 애플리케이션의 데이터 액세스가 데이터 객체를 명백히 명시하지 않고 대신에 요구된 객체의 애트리뷰트 상의 제약조건에 의해 논리적 도메인을 통해 이루어질 수 있는 능력. 이름 투명은 위치 투명과 중복 투명을 대체할 수 있음.
- 오너십(Ownership) & 비용(Costing) 투명 : 액세스 인증이나 자원사용비용의 측면에서 각 데이터 소스에 대한 액세스가 그 구조적 경계를 넘어 자유로울 수 있는 능력.
- 병렬(Parallelism) 투명 : 데이터를 처리하는 애플리케이션은 노드간의 병렬수행의 이익을 얻을 수 있어야 하며 자동적으로 노드를 추가(삭제)함으로써 확장(축)해야 할 수 있는 능력.
- 분산(Distribution) 투명 : 애플리케이션이 데이터를, 그 데이터가 분산된 소스의 셀로부터 얻게 됨을 모르 고서도 질의하고 갱신할 수 있는 능력. 또한 애플리케이션은 분산된 데이터를 통일된 방법으로 관리할 수 있어야 함.
- 위치(Location) 투명 (또는 네트워크 투명) : 애플리케이션이 데이터를 그 장소에 무관하게 액세스할 수 있는 능력.
- 중복(Replication) 투명 : 데이터를 액세스하는 애플리케이션은 성능 향상과 가용성을 위해 여러 장소에 중복되거나 캐쉬된 레플리카를 모르고서 그것들이 주는 이익을 취할 수 있는 능력.
- 분할(Fragmentation) 투명 : 데이터가 파일 전체가 아닌 파편으로 나뉘어 저장되어 있을 때 애플리케이션은 데이터의 어느 파편으로부터 데이터를 가져오느냐에 무관하게 액세스할 수 있는 능력.

2.2 데이터 연합에 의한 방법

정보 통합에 관한 연구는 연합 DBMS(federated DBMS) 분야에서 많이 수행되어 왔다.[7] 연합 DBMS란 분산된 환경에서 이질적인 데이터들을 액세스하도록 해주는 시스템이다. 연합 데이터베이스를 구축하는 방법에는 크게 전역스키마를 이용하는 방법과 래퍼(wrapper)를 이용하는 방법으로 나누어 볼 수 있다. 전역스키마를 이용하는 방법은 전역스키마라고 하는 통합스키마(global schema, universal schema)를 사용하는 것이다. 전역스키마를 사용하는 시스템에서는 스키마 변환 및 스키마 통합을 통해 전역스키마를 구성한 다음 이 전역스키마에 해당되는 데이터 조작언어를 이용하여 질의어를 표현한다. 이 기법을 이용하면 여러 데이터베이스 스키마를 통합하여 단일의 인터페이스를 제공하게 되므로 위치 투명(location transparency)을 얻을 수 있으나 스키마를 통합해야 하는 어려움이 너무 커서 대규모 정보통합에는 현실적이지 못하다. 이질적인 데이터베이스들을 연합하는 또 하나의 다른 방법인 래퍼를 이용하는 방법은 전역스키마를 사용하지 않는 대신 래퍼(wrapper)[15]라는 소프트웨어 모듈을 이용하는 것으로 사용자 질의에 대한 언어번역과 결과통합을 수행한다. 사용자 질의가 들어오면 연합 DBMS는 관련 데이터 소스에 대한 질의수행계획에 따라 서브 질의어로 바뀌어 각 데이터 소스에 대해 질의를 수행하게 된다. 래퍼를 이용한 방법은 과거 멀티 데이터베이스 언어 (multidatabase language)와 같은 강력한 데이터베이스 조작 언어를 이용하는 것과 비슷하다. 이러한 언어는 일반 데이터베이스 언어의 능력도 가지고 있어야 함은 물론 관련된 데이터베이스들이 상호작용(interoperable)할 수 있도록 지원해야만 한다. 여기서 상호작용이란 전역스키마의 부재와 자치성의 제약에도 불구하고 비절차적 오퍼레이션을 통해서 데이터베이스들을 사용할 수 있음을 의미한다.

정보통합의 수준을 분석해볼 때 래퍼를 이용한 방법은 앞에서 설명한 투명중 이질투명은 잘 제공되고, 분산투명은 비교적 잘 지원되나, 위치투명, 중복투명, 오너십투명 등은 제공되지 못한다.

2.3 데이터 그리드에 의한 방법

정보통합을 이루기 위한 다른 접근방법으로서 데이터 그리드(Data Grid)를 이용한 방법이 있다. 데이터 그리드는 그리드의 개념[9]을 데이터 부문에 적용한 것으로 이는 대용량의 데이터를 네트워크를 통해 공유하기 위한 접근이다. 그리드는 분산된 다수의 정보자원(컴퓨터, 데이터베이스, 측정장

치, 회의실 등)을 많은 사람들에게 공유할 수 있는 방법을 제공한다. 따라서 그리드는 전기, 수도, 가스 등과 같이 정보 플러그에 접속하기만 하면 언제 어디서나 필요한 때 필요한 만큼의 정보자원을 이용할 수 있는 새로운 사회 인프라로 기대되고 있으며, 그리드의 실현을 위하여 그리드 기술의 연구개발과 표준화 정책이 병행하여 진행되고 있다.

데이터 그리드란 개념은 2000년 Chervenak의 "The Data Grid"라는 논문[4]에 의해 그리드의 특수화 또는 확장으로서 소개되었다. 데이터 그리드는 대규모 데이터 소스들을 결합하고자 하는 요구를 만족시키기 위해 설계된다.[19] 이를 이용하면 여러 데이터 소스로부터 정보를 묶는 애플리케이션에 많은 진전이 있게 된다. 데이터 연합과 비교해볼 때 데이터 연합은 이미 상당부분 연구되었으나 데이터 그리드 애플리케이션의 요구를 잘 만족하지는 못하는 것 같다. 데이터 그리드는 다음과 같은 측면에서 데이터 연합과 차별된다.

- 데이터 그리드는 데이터 연합에 비해 메타데이터를 이용하여 데이터 소스를 매우 동적으로 선택 가능
- 데이터 그리드는 분산된 대규모 데이터 소스에 대해 큰 CPU와 네트워크로 인한 최고 성능 제공 가능
- 데이터 그리드는 데이터 복제로 인한 대체 데이터 소스 선택 가능

즉, 데이터 그리드와 데이터 연합과의 차이는 얼마나 높은 수준의 투명성을 제공하느냐로 생각해 볼 수 있다. 정보통합의 수준 측면에서 분석해보면 데이터 그리드는 이질투명, 분산투명, 위치투명, 중복투명, 오너십투명 등을 잘 지원한다.

III. 메타데이터 기반의 데이터 그리드

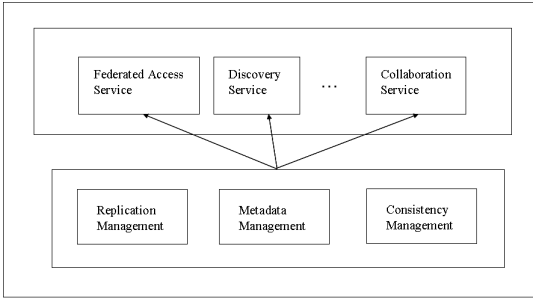
이 장에서는 데이터 그리드의 개념과 메타데이터에 관하여 기술한다.

3.1 데이터 그리드 개념

그리드는 계산 그리드와 데이터 그리드로 나눌 수 있다. 계산 그리드란 하드웨어 같은 컴퓨팅 자원을 공유하는 것을 다루고 데이터 그리드는 구조적으로 서로 다르고 멀리 떨어진 이질적인 데이터에 대한 유니폼하고 장소 무관한 액세스를 다룬다.

〈그림 3-1〉은 이러한 데이터 그리드의 개념을 보여 준다. 이 그림에서 아래 사각형은 데이터 그리드가 소요로 하는 필

수 기술을 보여주고 위의 사각형은 데이터 그리드가 제공하는 핵심 서비스를 보여준다. 주요 서비스를 설명하면 다음과 같다.[3.16]



〈그림 3-1〉 데이터 그리드 개념도
 〈Figure 3-1〉 Data Grid Concept Diagram

■ 디스커버리 서비스

디스커버리 서비스는 그리드 상에서 데이터의 이름과 위치 인식을 가능케 해준다. 애플리케이션은 데이터의 실제 위치보다는 논리적 도메인에 의해 데이터 소스를 명세한다. 이러한 논리적 도메인을 포함하는 요구를 받은 시스템은 디스커버리 서비스를 사용하여 이 논리적 도메인을 실제 데이터 소스 위치로 매핑한다.

■ 연합 액세스 서비스

연합 액세스 서비스는 그리드 소스의 분산 특성을 가상화 하는데 필수적이다. 연합 액세스 서비스의 주요 장점은 애플리케이션에게 하나의 데이터베이스라는 가상 이미지를 제공한다는 점이다. 이 가상의 하나의 데이터베이스는 전체로서 질의되고 갱신된다. 연합 액세스 서비스와 디스커버리 서비스의 결합은 애플리케이션에게 이질 투명, 분산 투명, 그리고 위치 투명성을 제공한다.

■ 협동 서비스

협동서비스는 모든 사용자에게 갱신을 전파하고 충돌을 해결하기 위하여 필요하다. 이 서비스의 주 역할은 독립적이고 분산된 데이터 갱신을 가상화하는 것이다.

3.2 메타데이터 기반 데이터 그리드

데이터 그리드를 이루는 기술에는 여러 가지가 있으나 여기서서는 그 중 가장 핵심이 되는 메타데이터를 중심으로 다룬다.

3.2.1 메타데이터와 데이터 그리드

메타데이터란 데이터에 관한 데이터이다. 좀 더 형식적으

로 살펴보면, 메타데이터는 메타와 데이터라는 말로 나뉜다. 옥스포드 사전에서 메타란 “더 높거나 2차 종류의 어떤 것”을 의미한다. 따라서 메타데이터는 문서에서 데이터에 있는 정보를 더욱 더 정교하게 만들기 위한 추가적인 데이터 층이다. 메타데이터는 최근 들어 데이터 응용과 관련된 여러 분야에서 사용되고 있다.[18] 메타데이터는 본질적으로 다이나믹하고 그 일생이 다양하므로 항상 최신의 내용으로 갱신되어야 하고 비용 효과적으로 관리되어야 한다.

정보통합에 있어서 직면한 핵심 이슈중 하나는 통합될 데이터를 위치시키고 이해하는 것이다. 메타데이터는 이러한 이슈에 직접적으로 사용될 수 있다. 그리드 기반 정보통합에 있어서 메타데이터의 핵심적인 역할은 다음과 같다.[6, 12]

- 저장 : 그리드를 구성하는 데이터를 쉽게 액세스가 가능하도록 저장한다.
- 데이터 발견 : 그리드 환경에서 관련된 데이터 소스를 찾고 액세스를 가능하게 하기 위한 스키마를 캡처하여 사용자에게 편리한 형태로 만들어주고 질의를 위한 소스로 매핑시켜 준다.
- 버전 및 형상 관리 : 스키마 진화를 관리하기 위해 스키마 버전과 그 매핑의 버전을 유지한다.
- 매핑 개발 : 그리드 구축을 위해 스키마정의 및 스키마 제약조건에 근거해서 스키마간의 매핑을 설계하고 매핑을 수행하기 위한 코드를 생성한다.

3.2.2 데이터 그리드를 위한 메타데이터 서비스

메타데이터는 데이터 그리드를 이루기 위한 핵심 요소기술이다.[5.6.13] 왜냐하면 메타데이터는 논리적 이름을 물리적인 데이터 이름으로 바꾸는 매핑에 사용되기 때문이다. 이러한 매핑을 위해 메타데이터를 저장하고 관리하기 위한 방안에는 메타데이터 레지스트리[20], 메타데이터 레퍼지토리[14], 메타데이터 카탈로그[17] 등 여러 방법이 있으나 여러 종류의 메타데이터를 공포하고, 저장하고, 액세스하기 위한 통일된 수단을 제공하는 메타데이터 서비스(service)를 제공한다. 이는 기본 개념에서 볼 때는 큰 차이가 없다.

데이터 그리드에 있어서 애플리케이션들은 메타데이터 서비스로 질의를 함으로써 관심있는 객체들을 식별한다. 이때 각 질의는 요구된 데이터의 특성을 명시해야 한다. 메타데이터 서비스는 그러한 특성과 전역적으로 유일한 이름을 갖는 엔티티인 논리적 객체를 연계시킨다. 이 논리적 객체는 물리적 인스턴스로는 하나 또는 그 이상의 데이터베이스 레코드 또는 파일일 수도 있다. 데이터 그리드 구축에 있어서 메타데이터 서비스는 데이터의 처리, 검색, 분석 및 해석을 지원하

고 또한 데이터의 오너십, 유효성, 그리고 질(quality)을 확립하는데 도움을 주어야 한다.[5] 또한 데이터 그리드 환경에서의 메타데이터 서비스는 메타데이터 저장장소와 표현에 대한 서로 다른 접근 방법을 통합하는 수단을 제공하는 것 외에도 분산 환경에서 효과적으로 작동해야 한다. 이는 확장가능(scalable)해야 하고 고장에 강해야 한다. 이러한 사항들을 고려해볼 때 데이터 그리드를 위한 메타데이터 서비스는 계층적이고 분산된 시스템으로서 구성되는 것이 바람직하다고 할 수 있다.

IV. 군사자료 데이터 그리드 시스템 구축

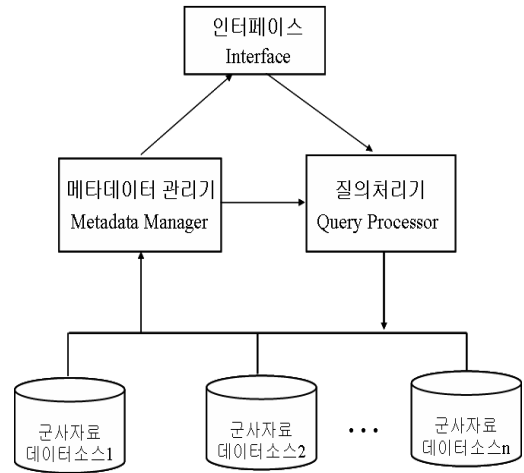
군사자료는 통상 단행본, 사진, 연감 등의 참고도서, 연구보고서, 석/박사 학위논문, 교범, 그리고 지도/필름, 동영상 등의 시청각 자료로 구성되어 있다. 본 논문에서는 이중 연구보고서, 논문 및 단행본을 위주로 앞에서 다룬 메타데이터와 데이터 그리드를 적용함으로써 군사자료를 통합 운용할 수 있는 군사자료 데이터 그리드 시스템을 제시한다.

4.1 시스템 개요

군사자료용 데이터 그리드 시스템은 <그림 4-1>에서 보는 바와 같이 질의처리에 관한 모든 작업을 담당하는 질의처리기(query processor)와 메타데이터 관리에 대한 모든 작업을 담당하는 메타데이터 관리기(metadata manager)로 구성된다. 질의처리기는 그리드 시스템의 핵심 역할을 수행하는 모듈중 하나로서, 사용자 인터페이스를 통하여 작성된 전역 질의를 받아 질의 분석과 질의 분해 및 질의 변환 작업을, 메타데이터 관리기의 도움을 얻어, 수행한다. 메타데이터 관리기는 사용자로 하여금 데이터의 물리적인 장소를 알지 않고서도 그 설명적 애트리뷰트의 값에 근거해서 데이터 셀을 찾아내도록 해준다. 따라서 장소-무관한 질의를 수행할 수 있다. 즉 메타데이터 애트리뷰트에 대해 질의를 하면 이러한 애트리뷰트를 포함하고 있는 여러 사이트에 분산되어 있는 모든 논리적 파일들을 검색하여 이들로부터 미리 정의된 매핑에 의해 원하는 정보를 검색해 준다. 따라서, 메타데이터 관리기를 이용하면 온톨로지를 이용하지 않고서도 전역 질의를 작성할 수가 있다. 이제 메타데이터 관리기와 질의처리기를 상세히 살펴본다.

4.2 군사자료 메타데이터 관리기

메타데이터 관리기는 앞 절에서 설명한 바와 같이 사용자로 하여금 데이터의 물리적인 장소를 알지 않고서도 그 설명적 애트리뷰트의 값에 근거해서 데이터 셀을 찾아내도록 해준다. 군사자료 메타데이터 관리기는 군사자료에 대한 메타데이터 서비스를 제공 관리하기 위한 시스템 모듈로써 군 특성을 고려할 때 그 요구사항은 다음과 같다.



<그림 4-1> 군사자료 데이터 그리드 시스템 구성도 개요
(Figure 4-1) Overview of Military Data Grid System

- 군사자료 메타데이터 관리기는 군사자료에 대한 효과적인 메타데이터 서비스를 지원하기 위해 집합체의 속성과 논리적인 이름의 속성을 연관시켜서 메타데이터 매핑을 모으는 능력을 뒷받침할 수 있어야 한다.
- 군사자료 메타데이터 관리기는 좋은 성능과 확장성을 가져야 한다. 군사자료 애플리케이션의 대부분은 실시간적이다. 실시간 군사 애플리케이션을 위해서는 지연시간이 짧아야 하고 높은 성능을 지원할 수 있어야 한다.
- 군사자료 메타데이터 관리기는 일관성, 인증, 권한부여, 그리고 감사여부와 관련한 확고한 정책을 포함해야 한다. 이는 군사 기밀자료에 대한 정확성 유지에 매우 중요한 요소이다.

4.2.1 군사자료 메타데이터 요소 설계

군사자료 메타데이터 관리기에서 다루게 되는 군사자료 메타데이터는 가장 널리 알려진 국제 표준인 더블린코어[21] 요소 중에서 도출하여 사용한다. 메타데이터 도출시 군사자료의 자원식별 및 선택에 필수적인 요소와 불필요한 요소가 무

엇인가를 고려하고 부가요소는 없는지 판단한다. 부가요소란 군사자료 메타데이터 관리기를 위하여 부가적으로 필요한 메타데이터를 의미한다. 더블링크어 15개 요소중 필수요소로 선정된 요소와 부가요소 및 불필요한 요소는 <표 4-1>에서 보는 바와 같다.

<표 4-1> 메타데이터 요소 분석
<Table 4-1> Analysis of Metadata Elements

필수요소	<ul style="list-style-type: none"> Title (표제) : 생산자나 출판자에 의해 자원에 부여된 이름 Creator (제작자) : 자원을 만드는 것에 대한 일차적 책임을 지는 개체 Subject (주제) : 자원의 토픽 Publisher (발행처) : 자원이 가용하도록 만드는 책임이 있는 개체 Date (발행일자) : 자원의 라이프 사이클에서 이벤트와 연관된 시점 또는 일정시간 Type (자료유형) : 자원의 본질 또는 장르 Format(형식) : 자원의 파일 포맷, 물리적 매체, 또는 차원 Language (언어) : 자원의 언어 Source(출전) : 기술된 자원이 유도된 자원
부가요소	<ul style="list-style-type: none"> 부대분류 : 군별, 군, 군단, 또는 사단을 값으로 가짐 비밀등급 : 1급, 2급, 3급 등의 비밀등급을 나타냄 직능분야 : 인사, 작전, 군수 등의 직능을 나타냄 병과 : 보병, 포병, 기갑 등의 값을 가짐
불필요요소	<ul style="list-style-type: none"> Description (설명) : 자원의 어카운트 Identifier (식별자) : 주어진 상황에서 자원에 대한 모호하지 않은 참조 Contributors (기타 제작자) : 자원에 대한 기여를 해야 할 책임이 있는 개체 Relation(관련자료) : 연관된 자원 Coverage(내용범위) : 자원의 공간적 또는 시간적 주제로 자원의 공간적 적용성, 또는 그 자원이 관련되는 관할 Rights(이용조건) : 자원 내외간에 걸쳐 있는 권한에 관한 정보

4.2.2 디스커버리 서비스 알고리즘

메타데이터 관리기에서 수행하는 디스커버리 서비스는 결국 앞에서 설명한 바와 같이 데이터 소스와 메타데이터 간의 매핑에 의해 이루어진다. 본 논문에서 제안된 시스템에서의 디스커버리 알고리즘은 다음과 같은 네가지 매칭 기법을 이용하여 서비스를 수행한다.

- Exact 매칭 : 데이터 소스 컬럼의 이름과 메타데이터가 정확히 일치하는 경우
- Synonym 매칭 : 데이터 소스 컬럼의 이름과 메타데이터가 동의어인 경우
- 패턴 매칭 : 데이터 소스 컬럼의 이름의 일부분과 메

타데이터의 일부만이 일치하는 경우

- 시맨틱 유사성 : 데이터 소스 컬럼의 이름과 메타데이터가 개념적으로 유사한 경우

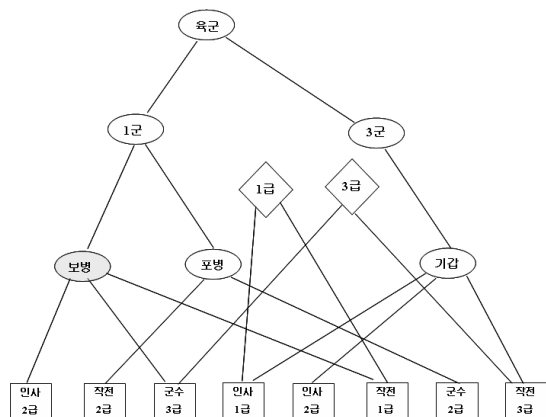
4.3 질의 처리기

질의처리기는 메타데이터 관리기의 도움을 받아 연합액세스 서비스를 제공한다. 이때 메타데이터 계층구조를 이용하여 질의 처리를 효과적으로 수행한다.

4.3.1 메타데이터 계층구조

효율적인 연합액세스 서비스를 위해 메타데이터 계층구조를 구축 이용한다. 메타데이터 계층구조(metadata hierarchy)란 메타데이터에 대한 계층구조[2]로써, 이는 메타데이터의 데이터 요소 간의 계층구조와 메타데이터의 도메인 값 간의 계층구조로 나누어 생각할 수 있다. 본 논문에서의 메타데이터 계층구조는 이 중 도메인 값 간의 계층구조를 이용하는 것으로 군사자료가 비밀등급과 부대분류 등에 대하여 포함관계를 이용한 계층구조를 갖는 것에 기초한다. 메타데이터 요소중 비밀등급을 나타내는 요소와 부대분류를 나타내는 요소에 대한 계층구조를 작성하면 직능과 병과정보와 연계되어 군사자료에 대한 그룹핑이 요구될 때 매우 유용하게 쓰일 수 있다. 메타데이터 계층구조는 효과적인 정보제공을 위하여 다음과 같은 논리적 구조를 지원한다.

- 집합(agggregation) : 원하는 단계별 정보 제공 기능으로 요구되는 정보를 AND하여 다양한 요구에 대응할 수 있음.
- 뷰(view) : 원하는 특정 정보를 AND/OR를 이용하여 구체적으로 지원



<그림 4-2> 메타데이터 계층구조
<Figure 4-2> Metadata Hierarchy

〈그림 4-2〉는 이러한 메타데이터 계층구조를 보여준다. 이 그림에서 사각형은 데이터 파일 객체(데이터베이스 레코드 일 수도 있음)를 나타내고 원은 수직적으로 집합화되는 객체를 나타낸다. 그리고 다이아몬드는 수평적으로 집합되는 객체의 보안등급을 나타내는 객체이다. 원과 다이아몬드 같은 집합객체는 AND 또는 OR를 이용하여 뷰를 구성할 수 있다. 다시 말해서 뷰는 집합화된 객체들을 AND나 OR로 묶어주는 역할을 담당한다.

4.3.2 질의 처리

〈그림 4-2〉에서 부대분류와 보안등급, 병과, 군사직능을 이용하여 계층구조가 만들어졌음을 알 수 있다. 계층구조는 사용자 인터페이스와 연계되어 사용자 질의 작성을 돕는다. 이 구조를 이용하면 이들 간에 그루핑된 정보를 용이하게 검색할 수 있다. 예를 들어, 1군 예하 보병에 관한 자료의 발행처와 날짜를 보려면 그림에서 음영처리된 “보병”을 검색하면 된다(Q₁참조). 또한 구체적인 정보는 뷰를 이용하여 접근 가능하다. 예를 들어, 1군에 속하면서 1급비밀인 자료의 발행처와 날짜는 1군과 1급의 andview로서 얻어질 수 있다(Q₂ 참조). 이러한 질의를 집합 기능이 추가된 SQL-like 질의로 나타내면 각각 Q₁ 및 Q₂와 같다.

Q₁ Select publisher, date
From Aggre_보병
Where 부대분류='1군'

Q₂ Select publisher, date
From Andview_1급
Where 부대분류='1군'

이러한 SQL-like 전역질의가 처리되는 절차는 〈그림 4-3〉에서 보는 바와 같이 메타데이터 관리기 및 래퍼의 도움을 얻어 해석되고 분해되어 각 사이트에서 해당 질의어로 실행되고 그 결과는 합성되어 사용자에게 전달된다. 예를 들어 질의 Q₁이 처리되는 알고리즘은 다음과 같다.

Procedure QueryProcessing

Step 1. 전역질의 Q₁이 해석된다.

{Publisher, Date}←{1군&인사, 1군&군수, 1군&작전}

Step 2. 메타데이터 관리기의 디스커버리 서비스에 의해서 위 자료가 각각 사이트 S₁, S₂, S₃에 위치하고 있는 XML DB의 인사파일, 관계형 DB의 교

범 테이블에, 객체지향 DB의 동계작전 테이블에 각각 있음이 파악된다.

{Publisher, Date}←{1군&인사&XMLDB}

{Publisher, Date}←{1군&군수&관계형DB}

{Publisher, Date}←{1군&작전&객체지향DB}

Step 3. 래퍼에 의해 각각 해당질의어인 XQuery, SQL, OQL로 각각 변환 수행된다. 예를 들어 SQL로의 변환은 다음과 같다.

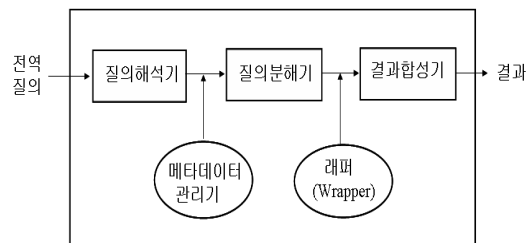
Select Publisher, Date

From 교범

Where 부대분류='1군' and 직능분야='군수'

Step 4. 각 사이트에서 수행된 결과가 합성되어 사용자에게 제공된다.

End QueryProcessing



〈그림 4-3〉 질의 처리 절차
(Figure 4-3) Steps for Query Processing

V. 결론

최근 들어 유비쿼터스 시대를 맞이하여 정보의 활용이 점점 더 부각되는 이 시대에 정보의 통합은 매우 중요하다. 정보통합을 이루기 위한 한 방법으로서 데이터 그리드를 이용한 방법이 대두되고 있다. 데이터 그리드란 그리드의 개념을 데이터 부분에 적용한 것으로, 구조적으로 서로 이질적이고 멀리 떨어진 대규모의 데이터에 대한 공유를 위해 유니폼하고 장소 무관한 액세스를 가능하게 해 준다.

본 논문에서는 군 정보통합을 위한 접근방법으로서 메타데이터 기반의 군사자료 데이터 그리드 시스템을 제시하였다. 메타데이터 관리기는 사용자로 하여금 데이터의 물리적인 장소를 모르더라도 그 설명적 애트리뷰트의 값에 근거해서 데이터 셀을 찾아내도록 해준다. 따라서 장소-무관한 질의를 수행할 수 있다. 질의처리기는 효율적인 연합액세스 서비스를 위

해 메타데이터 계층구조를 구축 적용하였다. 따라서 제안된 시스템은 그리드를 위한 핵심 투명인 위치투명, 이질 투명 및 분산투명을 지원하므로 그리드 기능을 충분히 제공할 수가 있다. 따라서 본 연구 결과는 지식관리형 국방 정보통합시스템 구축, 효율적인 국방 정보검색시스템 구축, 또는 국방정보체계통합 등 여러 분야에 기반기술로서 활용될 수 있을 것이다. 이와 같은 기반기술의 확보는 군 예산절감에도 기여할 수 있을 것으로 생각된다.

앞으로의 연구로서는 메타데이터 계층구조에 대하여 좀 더 심도 깊은 연구가 필요하고, 군사자료 질의처리 알고리즘에 관한 연구도 계속되어야 할 것이다. 또한 최근 미군에서 NCW(Network Centric Warfare) 환경을 지원하기 위해 제시한 정보그리드[8]와 본 연구와의 연계성도 연구되어야 할 것이다.

참고문헌

- [1] S. Amer-Yahia, P. Case, T. Rolleke, et al.(2005), "Report on the DB/IR Panel at SIGMOD 2005", SIGMOD Record, Vol 34, No. 4, pp.71-74.
- [2] A. Assali, and H. Zanghi(2006), "Automated Metadata Hierarchy Derivation", Information & Communication Technologies: from Theory to Applications (ICTTA'06).
- [3] M. Atkinson, et al., "Grid Database Access and Integration: Requirements and Functionalities", www.ggf.org/documents/GFD.13
- [4] A. Chervenak, I. Foster, C. Kesselman, et al.(2000), "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets," Journal of network and Computer Applications, 23, pp.187-200.
- [5] O. Corcho, P. Alper, P. Missier, S. Bechhofer, C. Oble(2007), "Grid Metadata Management: Requirements and Architecture," The 8th IEEE/ACM International Conference on Grid Computing, pp.97-104.
- [6] E. Deelman, G. Singh, M.P. Atkinson, et al.(2004), "Grid-based Metadata Services," SSDBM04.
- [7] R. Elmasri and S. B. Navathe(2004), Fundamentals of Database Systems, 4th ed. Addison Wesley.
- [8] L. Ferreira, et al., "The Information grid, Part 1: The Infrastructure", www.ibm.com/developerworks/library/gr-info1/
- [9] I. Foster, C. Kesselman, S. Tuecke(2001), "The Anatomy of the Grid", International J. of Supercomputer Applications, vol.15, pp.200-222.
- [10] A. Haalevy, N. Ashish, D. Bitten, et al.(2005), "Enterprise Information Integration: Successes, Challenges and Controversies", SIGMOD, pp.778-787.
- [11] A. Haalevy, A. Rajaraman, J. Ordille, et al.(2006), "Data Integration: The Teenage Years", VLDB, pp.9-16.
- [12] R. Hauch A.Miller, R. Cardwell(2005), "Information Intelligence : Metadata for Information Discovery, Access, and Integration", SIGMOD, pp.793-798.
- [13] K. Kaczmarek, P. Habela, K. Subieta(2004), "Metadata in a Grid Construction", IEEE WETICE'04.
- [14] M. Matula, "Netbeans Metadata repository", mdr.netbeans.org/MDR-whitepaper.pdf
- [15] E. Ozsu and Valduriez(1999), Principles of Distributed Database Systems(2nd), Prentice Hall.
- [16] V. Raman et al., "Services for data Access and data Processing on Grids", www.ggf.org/documents/GFD.14
- [17] G. Singh, S. Bharathi, A. Chervenak, E. Deelman, C. Kesselman, M. Manohar, S. Patil, L. Pearlman(2003), "A Metadata Catalog Service for Data Intensive Applications", SC 2003.
- [18] D. Srivastava, Y. Velegrakis(2007), "Intentional As sociations Between Data and Metadata", ACM SIGMOD, pp.401-412.
- [19] S. Venugopal, R. Buyya, K. Ramamohanarao(2006), "A taxonomy of Data Grids for distributed data sharing, management, and processing", ACM Computing Surveys, Volume 38 Issue 1.
- [20] ISO/IEC 11179-1(2004) Information technology - Metadata Registries(MDR) - Part 1 : Framework.
- [21] dublincore.org/documents/dces

저 자 소개



나 민 영

- 육군사관학교 졸업
- 서울대학교 컴퓨터공학과 졸업
- 서울대학교 대학원
컴퓨터공학과 졸업, M.S.
- University of Florida, Dept.
of Computer and Information
Sciences, Ph.D.
- 미 IBM Watson 연구소 연구원
- Georgia Institute of Technology
교환교수
- 현재: 육군사관학교 전자정보학과 교수