

# 컴포넌트 메타데이터의 정의를 통한 효율적인 검색 방법에 관한 연구

\*김행곤 차정은 김병준<sup>U</sup> \*\*김철홍

\*대구가톨릭대학교 컴퓨터공학과 \*\*ETRI 컴퓨터소프트웨어 공학연구부  
{hangkon, jecha}@cuth.cataegu.ac.kr, windjun@kebi.com \*\*kch@etri.re.kr

## A Study for Effective Retrieval Method Through Definition of Component Metadata

\*Heang-Kon Kim Jung-Eun Cha Byung-Jun Kim<sup>U</sup> \*\*Chul-Hong Kim  
\*Dept. of Computer Engineering, Catholic University of Daegu  
\*\*ETRI Computer & Software Technology Lab Se Department

### 요 약

소프트웨어 공학의 새로운 패러다임인 컴포넌트가 지니는 다양한 속성들과 장점은 아직 많은 문제점들이 남아있는 상태이지만 소프트웨어 개발시의 높은 생산성과 효율성, 그리고 재사용성과 유지보수의 이점으로 크게 주목받고 있다. 컴포넌트의 뛰어난 장점들을 효과적으로 이용하기 위해선 우선 표준화된 명세를 통해 컴포넌트가 개발되어 유통 업체나 벤더의 저장소에 저장되어 있어야 하며 구축된 저장소에서 사용자는 컴포넌트 검색 서비스를 이용하여 자신의 요구에 적합한 컴포넌트를 구매, 사용할 수 있는 체계화된 프로세스가 필요하다. 따라서 본 논문에서는 컴포넌트 검색을 위한 선행 작업으로서, 검색용 메타데이터를 메타데이터가 가져야 하는 원칙을 적용하여 정의한다. 컴포넌트가 가지는 일반적인 속성과 성질로부터 사용자가 컴포넌트를 검색하고자 할 때 사용되어질 수 있는 키워드를 메타데이터의 항목으로 추출하였으며 이렇게 정의된 메타데이터를 바탕으로 일반적인 정보의 분류에 적합한 기존 웹 검색과는 다른 컴포넌트만의 검색을 다루는 새로운 검색 시스템을 설계하였다. 따라서 효과적인 컴포넌트 재사용을 위해 검색용 메타데이터와 이를 바탕으로 한 컴포넌트 검색 시스템을 제안하고자 한다.

### 1. 서론

웹 상에서의 효과적인 정보 검색의 관건은 사용자가 요구하는 정보의 키워드를 어딘가에 존재하고 있을 정보에 정확하게 매핑 시켜주는 검색엔진 또는 검색서비스에 있다. 같은 맥락으로 CBD(Component Based Development)의 가치 있는 부산물들의 효과적인 사용을 위해선 컴포넌트 사용자의 요구가 담긴 키워드와 컴포넌트의 개발의도나 응용범위를 정확히 매핑시킬 수 있는 검색용 메타정보 또는 검색 시나리오가 필요하다[3].

최근 들어 연구, 정착되고 있는 CBD 방법론은 폭발적인 소프트웨어 수요와 다양한 비즈니스 현장의 빠른 요구 변화, 그리고 e-Business로의 추세를 반영하고 있다. 따라서 세계 곳곳에서 많은 양의 컴포넌트들의 개발이 이루어지고 있으며 빠른 몇몇 업체들에 의해 이미 유통되고 있다. 또한 개발기간의 단축이나 개발비용감소를 목적으로 e-Business를 위한 컴포넌트를 찾는 일반개발자들의 요구가 증가하고 있다. 하지만, 이러한 흐름에도 불구하고 컴포넌트의 성과는 아직 그 기대와 노력에 미치지 못하고 있다. 그 이유로는 몇 가지 꼽을 수 있는데 첫 번째, 컴포넌트 개발과 컴포넌트 유통에서의 표준명세의 부재이다. 이러한 이유로 컴포넌트 개발자의 정확한 의도파악이 어려우며 컴포넌트 본래의 기능성에 신뢰감이 줄어든다. 또 하나는 컴포넌트 유통에서 기존에 존재하던 키워드 검색이나 브라우징에 의존한 컴포넌트의 검색의 문제점이다.

컴포넌트는 웹에서의 문서나 이미지로만 존재하던 기존 정보들의 형태와는 달리 이진형태로 되어 있어 잘 정리된 검색을 위한 컴포넌트 메타데이터가 정의되어야만 하며 기존의 정보 검색과는 다른 새로운 검색 메커니즘이 제시되어야만 할 것이다. 본 논문에서는 컴포넌트의 환경적, 기능적, 서술적 속성을 고려, 컴포넌트 검색을 위한 메타데이터를 정의하고 이를 참조하는 검색 시스템을 제안한다.

### 2. 관련연구

#### 2.1 컴포넌트 메타데이터

메타데이터(MetaData)는 데이터가 가지는 속성과 구조, 물리적인 위치의 변화 이력, 연결성 등 하나의 데이터가 가지는 가장 기본적인 사항을 포함하고 있는 데이터의 데이터를 말한다. 또한 메타데이터는 애플리케이션 시스템에서의 저장은 물론 검색, 관리, 운영, 유지보수 등에 결정적인 역할을 수행한다. 메타데이터는 각종 데이터 즉, 문서, 파일, 미디어 등을 효과적으로 운영 관리할 수 있게끔 하여 정보의 고효율, 시스템의 용이한 유지보수를 가능케 한다. 따라서 데이터의 메타데이터는 저장소와 상호 밀접하게 결합돼 있다. 결과적으로, 사용하고자 하는 정보의 메타데이터 정의는 시스템의 원활한 운영과 초기의 목적 달성을 위해 필수적이며 신중하여야 할 것이다. 본 논문에서는 국제협회 권고안과의 비교를 통해 컴포넌트의 검색을 위한 메타 데이터를 정의하고자 한다[5].

본 논문은 2000년도 한국전통산업연구원 컴퍼넌트 사업 계획 지원에 의해 수행됨

<표 1> CEO의 메타데이터 권고안

title	supply details	organization
abstract	publisher	point of contact
language	publication date	author
subject index	record identifier	spatial coverage
temporal coverage	creation date	cross references
resource identifie	expiry date	scale
resolution	resource type	quality
supply details	provider	constraints
last modification	version	

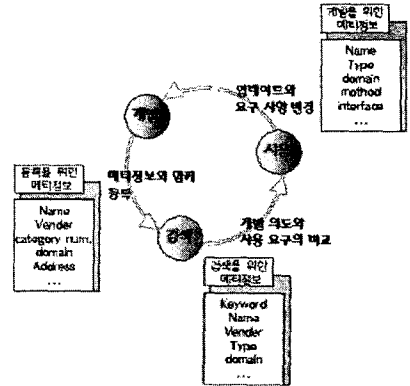
<표 1>은 CEO(Center for Earth)에서 권고하는 메타데이터이다. 또한 인터넷과 네트워크에 사용되어지는 더블린 코어는 자원 기술을 목적으로 데이터의 기본적인 요소의 선정과 기술적인 기준을 다음과 같이 제시하였다. 자원의 본질적인 특성을 기술요소로 하되(고유성), 규정된 필수데이터요소 이외에 부차적인 내용이나 특성을 기술요소로 사용하며(확장성), 응용분야나 표현기법을 규정하지 않으며(구문의 독립성), 각 요소의 수록 여부를 강제하지 않으며(선택성), 모든 기술요소는 반복 사용할 수 있으며(반복성), 한정어를 사용하여 세부사항을 조정(수정가능성)한다는 것이다. 이것은 보편적인 웹기반 환경과 같이 네트워크환경에서 사용되는 데이터의 표준형식이 이식성, 내구성, 호환성, 확장성, 융통성 등을 확보하는 것이 중요하기 때문이다[5].

2.2 컴포넌트 유통에서의 검색

현재 웹상의 컴포넌트 검색을 살펴보면 flashline(www.flashline.com)의 경우 Type(Java Product, Com Product)과 vender에 따른 일차적인 분류 후 브라우징하는 검색 방법과 키워드 매핑에 의한 검색을 제공하고 있다. component source(www.componentsource.com)의 경우 카테고리를 중심으로 검색하며 component registry(www.componentregistry.com)는 class name과 keyword 의한 검색을 제공한다. 일반적으로 브라우징 기법은 컴포넌트를 몇 가지 기준에 의해 분류해 놓은 것으로 찾고자 하는 컴포넌트 검색을 시각화한다는 점에서 '체감 정확도'는 상승하지만 실제 사용자가 원하는 컴포넌트를 찾기 어렵다.[4]

3. 컴포넌트 검색용 메타 데이터의 정의

컴포넌트의 메타정보는 사용자의 목적과 사용 범주에 따라 달라질 수 있다. 따라서 개발에 필요한 메타정보, 사용에 필요한 메타정보, 그리고 사용에 앞서 사용자의 요구를 정확히 받아들일 수 있도록 검색을 위한 메타데이터로 구분하여 생각해볼 수 있다. (그림 1)는 컴포넌트 기반 개발의 각 단계마다 요구되고 참조되는 메타데이터의 다양함을 나타내고 있다. 본 논문에서 제시하는 컴포넌트 검색을 위한 메타데이터는 모든 정보로서의 단위가 가지는 형식적 속성과 컴포넌트가 가지는 특성인 기능적 속성, 그리고 환경적인 속성을 검색 메타 데이터로서 제시함으로써 컴포넌트만의 효율적인 검색을 이루고자 한다. 또한 보편화 될 컴포넌트 유통에서 수요자가 필요한 정보의 체계를 구축하고자 한다[1].



(그림 1) CBD 각 단계에서의 메타 데이터

본 논문에서는 컴포넌트 검색을 위한 메타데이터를 다음의 <표 3>과 같은 원칙에 입각, 정의하였다.

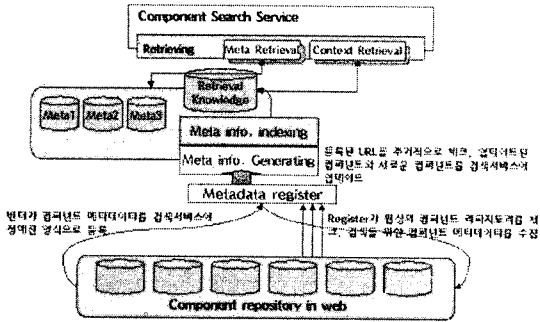
<표 3> 메타 데이터 정의 원칙

- ① 공인된 표준메타형식에 기반 한다. (일반성)
- ② 단순하면서도 구조화된 형식을 갖는다. (단순성, 형식성)
- ③ 비즈니스 도메인 선정에 있어 카테고리 되어진 아키텍처 기반의 분류정보를 사용한다.
- ④ 컴포넌트의 기능적 속성을 표현한다. (기능성)
- ⑤ 메타데이터 요소는 함축적 의미를 내포한다. (함축성)

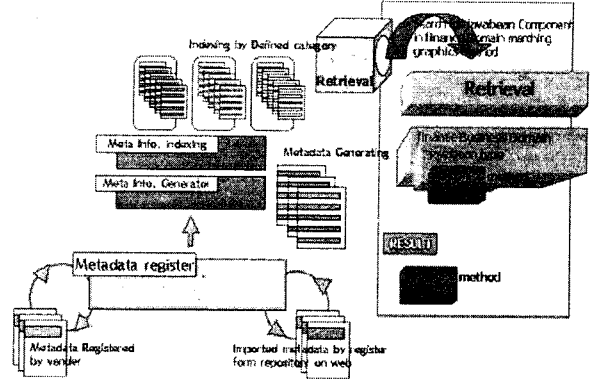
컴포넌트만의 기능적인 속성들을 위한 항목의 하나인 'element : method' 요소의 선정으로 개발자 정의 메소드와 구현 언어 종속적인 메소드를 참조, 요구되는 컴포넌트의 기능성과 적용 가능성을 사용자가 쉽게 이해할 수 있도록 했다. 또 하나의 기능 속성인 'element : interface'를 선정함으로써 다른 컴포넌트와의 연계성을 고려했다. 다음의 <표2>은 컴포넌트의 여러 가지 일반적인 속성들로부터 검색을 위해 추출한 메타데이터이다.

<표 2> 검색을 위한 메타 데이터

General Property	
name	component name
vender	developer or provider
business domain	Problem which can be applied
container	Environment and Type
price	cost
location	located URL
Functional Property	
method	user defined method or specific method
interface	interface of component
Descriptive Property	
abstract	summary
creation date	last modification



(그림 2) 검색 시스템 전체 구성도



(그림 3) 메타데이터의 수집과 분류, 검색

#### 4. 컴포넌트 메타정보를 이용한 검색 시스템

##### 4. 1 시스템 구성

본 논문에서 제시하는 컴포넌트 검색 시스템(그림 2)은 컴포넌트 검색을 위한 메타데이터 구축을 위해 두 가지 선행 작업을 수행한다. 컴포넌트 개발자, 벤더, 중소기업된 유통업체가 새로운 컴포넌트 개발, 확장시 검색 시스템의 정해진 양식으로 메타데이터 레지스터에 등록하는 첫 번째 방식과 메타데이터 레지스터가 웹상의 컴포넌트 유통업체를 방문, 체크하여 외부 컴포넌트 검색을 위한 자료를 수집해온다. 수집된 자료는 메타 데이터 제네레이터에 의해 검색용 메타 데이터로 생성되며 검색시스템에 저장된다. 메타 데이터 저장시 비즈니스 도메인을 기준, 관련성이 높은 데이터들끼리 자동 분류함으로써 검색시의 성능을 높일 수 있다. 이렇게 카테고리 되어진 메타데이터는 컴포넌트 정보의 업데이트에 대비, 메타데이터 레지스터가 이미 등록된 URL을 기준으로 주기적으로 체크하여 메타데이터 제네레이터에 의해 메타데이터를 재작성하며 변경 사항이 비즈니스 도메인과 관계가 있을 시 재분류되어 검색을 위한 최선의 데이터를 보유하게 된다.

##### 4. 2 컴포넌트 검색

컴포넌트 검색을 위해 사용자는 기본적인 질의를 입력하며 메타데이터 요소의 타입에 따라 선택적으로 더 많은 부가 사항을 지정할 수 있다. 사용자는 먼저 비즈니스 도메인(finance Business Domain)을 정한다. 원하는 컴포넌트의 타입(javabean Component)을 지정하고, 부가적으로 컴포넌트의 기능적 속성인 method 요소(graphics Method)를 사용하거나, 정의된 메타데이터 요소를 적극 활용함으로써 사용자가 원하는 컴포넌트에 가장 부합되는 결과를 얻을 수 있도록 한다<표 4>.

<표 4> 검색 질의의 예

Search for **javabean** Component  
in **finance** Business Domain  
Marching **graphics** Method  
by **Sun** Company  
.....

본 논문에서 제안하는 컴포넌트 검색 시스템에서 사용자가 컴포넌트를 검색하는 과정을 (그림 3)에 도식화하였다. 사용자가 원하는 비즈니스 도메인의 컴포넌트부터 결과내 검색을 사용, 원하는 타입, 원하는 기능적 속성등 단계, 단계, 반복적인 검색을 수행하여 결과로 검색되어진 컴포넌트를 메타데이터 요소 중 'abstract' 항목과 'URL' 항목과 함께 제공한다. 사용자의 부가적인 검색 옵션이 많을수록 컴포넌트 검색의 '정확율(Precision)'은 상승하며 적을수록 관련된 컴포넌트 검색 결과의 '재현율(Recall)'이 높아질 수 있다.

#### 5. 결론 및 향후 연구

본 논문에서는 일반적인 정보와 차별된 속성을 가지는 컴포넌트를 효율적으로 검색하기 위한 선행 작업으로서 컴포넌트가 가지는 다양한 속성들로부터 검색을 필요한 속성들을 추출, 메타데이터로서 정의하였으며, 이 메타데이터를 참조하는 컴포넌트 검색 시스템을 제시하였다. 향후 연구로는 메타정보 요소의 표준화와 컴포넌트 표현의 표준포맷의 정의, 그 의미의 확장을 바탕으로 컴포넌트 검색과 사용의 질을 앞당겨야 할 것이다.

#### 6. 참고 문헌

- [1]. CBOP, 메타데이터 구조 표준 V.01, CBOP 리퍼지토리 메타 데이터 구조 정의서, 1999
- [2]. Clemens Szyperski, Component Software, Addison-Wesley, 1998
- [3]. Robert C. Seacord, et al "Agora : A Search Engine for Software Components", Carnegie Mellon Technical Report
- [4]. 마이크로 소프트웨어 2000. 3, 박민우의 4명 "인터넷 세상, 검색 엔진이 주도한다"
- [5]. 메타데이터 구조론, <http://dewey.yonsei.ac.kr/metadata/>