

## 온톨로지 기반 영화 메타데이터간 연관성을 활용한 영화 추천 기법

김재영  
아주대학교 컴퓨터공학과  
(vlqtc05@ajou.ac.kr)

이석원  
아주대학교 소프트웨어 융합학과  
(leesw@ajou.ac.kr)

최근 IPTV와 스마트 TV 등의 등장과 영상 콘텐츠를 시청하고 검색할 수 있는 웹 서비스의 등장으로 영상 콘텐츠의 접근이 용이해져 사용자들은 자신이 원하는 콘텐츠를 찾고자 하는 요구가 증가하고 있다. 하지만 서비스되는 콘텐츠의 양이 방대하여 영상 콘텐츠를 검색할 때 사용하는 키워드 기반의 검색은 많은 양의 결과를 가져오며 사용자가 필요로 하지 않는 결과가 검색된다. 따라서 사용자가 원하는 콘텐츠의 검색 시간과 노력이 증가 하게 되었다. 이를 극복 하기 위해 콘텐츠 추천 및 검색에 대한 연구가 수행되어 왔다. 기존의 연구에는 사용자의 선호도 분석을 통하여 영상 콘텐츠를 추천하거나 비슷한 성향을 가지는 사용자들을 분류하여 콘텐츠를 추천하는 기법들이 연구되어 왔다. 본 논문에서는 영상 콘텐츠 중 영화의 추천을 위해 사용자 개인의 영화 메타데이터의 선호도를 분석하고, 영화의 메타데이터와 영화의 유사성을 도출하여 이를 기반으로 영화 추천 기법을 제안한다. 영화의 특징을 담고 있고, 사용자의 영화 선호도에 영향을 끼치는 장르, 줄거리, 배우, 키워드 등의 영화 메타데이터를 기반으로 온톨로지를 구축하고, 확률 기법을 통한 메타 데이터간의 유사성을 분석하여 유사 메타데이터를 연결한다. 또한 사용자의 선호도와 그룹을 정의하고, 사용자 정보를 활용하기 위한 사용자 모델을 정의한다.

제안하는 추천 기법은 1) 사용자 정보기반의 후보 영화 검색 컴포넌트, 2) 사용자 선호기반의 후보 영화 검색 컴포넌트, 3) 1)과 2)의 결과를 통합하고 가중치를 부여하는 컴포넌트, 4) 최종결과 분석을 통한 개인화된 영화 추천 컴포넌트 등 총 4가지 컴포넌트로 구성된다. 제안하는 추천 기법의 실험을 위하여 20대 남/녀 10명씩 20명을 대상으로 실험을 진행하였으며, 실험결과 평균 Top-5에서 2.1개 Top-10에서 3.35개 Top-20에서 6.35의 영화가 보고 싶은 영화로 선택되었다.

본 논문에서는 영화 메타데이터간의 연관성 도출을 통하여 영화간의 유사성을 도출하고 이를 기반으로 사용자의 기본적인 정보를 활용한 추천뿐만 아니라 사용자가 예상하지 못한 영화의 추천이 가능하다.

논문접수일 : 2013년 06월 27일    논문수정일 : 2013년 08월 30일    게재확정일 : 2013년 09월 04일

투고유형 : 국문급행    교신저자 : 이석원

### 1. 서론

IPTV(Internet Protocol Television)와 스마트 TV의 등장과 영화나 TV 프로그램 같은 영상 콘텐츠를 검색하고 시청할 수 있는 웹 서비스의 등장으

로 사용자는 영상 콘텐츠에 접근이 용이해졌으며 기존의 영상 콘텐츠 시청 방법과 달리 시간에 구애 받지 않고 상영이 되었던 영화나 TV 프로그램의 검색과 시청이 가능하게 되었다. 이에 따라 원하는 영상 콘텐츠를 찾고자 하는 요구가 증가되었다. 하

\* 본 연구는 2012학년도 아주대학교 일반연구비 지원에 의하여 연구되었음.

지만 서비스되는 영상 콘텐츠의 양이 방대하여, 원하는 영상 콘텐츠를 검색하기 위해 많은 시간과 노력을 필요로 한다. 이러한 문제점을 극복하기 위해 영상 콘텐츠의 검색과 추천을 쉽게 제공할 수 있는 방법에 대한 다양한 연구가 진행되어 왔다(Noy et al., 2007; Yun et al., 2010; Kim et al., 2011).

일반적으로 추천 서비스에서는 제품을 판매하기 위해 사용자가 선호하는 제품과 제품의 정보를 제공한다. 추천시스템에 사용하는 대표적인 방법으로는 협업적 여과(Collaborative Filtering)와 내용기반 여과(Contents Based Filtering)가 있다. 협업적 여과는 사용자와 비슷한 선호도나 정보를 가지는 다른 사용자들을 찾고 그 사용자들이 과거에 구매했던 아이템이나 서비스의 이력을 기반으로 추천해 준다. 내용기반 여과는 사용자가 과거에 구매했던 아이템이나 서비스와 유사한 아이템이나 서비스를 사용자에게 추천해 준다. 이러한 추천기법들은 다양한 도메인에서 추천 서비스로 활용되고 있다(Adomavicius et al., 2005).

본 논문에서는 사용자의 프로파일과 영화의 메타데이터에 대한 선호도를 분석하여 영화를 검색하고 추천한다. 사용자가 영화를 선택할 때 고려하고 영화의 특징을 담고 있는 메타데이터를 기반으로 영화 메타데이터 온톨로지를 구축하여 이를 기반으로 영화를 검색한다. 기존의 영화의 정보를 제공하는 서비스에서는 단지 특정 영화의 메타데이터를 나열함으로써 영화의 정보를 제공한다. 본 논문에서는 영화의 메타데이터를 온톨로지 구조의 메타데이터 지식베이스로 구성하고 이를 사용하여 영화 메타데이터간의 속성을 부여함으로써 메타데이터간의 관계를 정의한다. 또한 영화 메타데이터간의 유사성 분석을 통하여 영화간의 유사성을 도출하고 검색된 영화의 메타데이터를 확장함으로써 다른 유사한 영화를 검색한다.

영화 메타데이터 온톨로지는 영화 메타데이터간의 관계와 속성을 생성하고, 생성된 관계와 속성을 통한 영화간의 유사성을 도출함으로써 유사한 영화추천에 활용될 수 있다. 영화간 유사성을 도출하기 위해 본 논문에서는 장르를 활용한다. 장르간의 연관성 도출을 위해 확률적 방법을 적용하고 이를 통해 정의된 영화 간 유사성 모델을 사용하여 제안하는 추천기법에 적용한다. 또한 개인화된 영화 검색을 위해 줄거리를 사용한다. 보통의 영화 검색 서비스에서 영화를 검색할 때 사용하는 메타데이터는 장르, 감독, 배우, 키워드 등이 있다. 이러한 메타데이터들은 각 메타데이터에 해당하는 영화를 대표하는 정보를 가지고 있다. 하지만 특정 한 영화의 정보를 제공해 주지는 못한다. 줄거리는 일반적으로 영화 검색에 사용되는 장르, 감독, 배우, 키워드와 같은 메타데이터와는 다르게 해당영화의 간단한 내용을 담고 있기 때문에 사용자에게 특정 영화를 대표하는 정보를 전달해 줄 수 있다. 따라서 영화 메타데이터 온톨로지 에서 줄거리 검색을 통하여 사용자가 원하는 영화를 검색하고 추천 할 수 있으며 영화 검색을 할 때 사용하는 대표 메타데이터들은 다르지만 유사한 내용을 담고 있는 영화의 검색이 가능하다. 따라서 영화와 같은 영상 콘텐츠 추천 기법에서 사용하는 선호 장르기반의 콘텐츠 추천 시스템과 달리(Kim et al., 2011; Chen et al., 2007) 유사한 줄거리를 가지지만 장르가 달라, 추천되지 못하는 경우를 방지할 수 있으며, 사용자가 선호하는 유사 영화를 추천함으로써 사용자의 만족도를 높일 수 있다.

또한, 사용자의 정보를 활용하여 사용자에게 적절한 영화를 추천해 줄 수 있도록 사용자 모델을 제안한다. 사용자 모델에는 개인이 선호하는 영화 메타데이터와 사용자의 성별과 연령 정보를 포함

한다. 사용자의 성별, 연령 같은 정보를 기반으로 사용자 모델에서 사용자 그룹을 결정한다. 본 논문에서는 성별, 연령 등의 정보를 활용하기 위하여 영화진흥위원회(Korean Film Council, 2011)에서 조사한 연령별, 성별 기반의 선호 장르를 기반으로 영화를 검색하고 추천한다. 그리고 사용자 개인의 선호도를 기반으로 각 메타데이터에 같은 가중치를 부여하여 영화를 검색하고 추천한다. 추천된 영화의 결과와 사용자가 선택한 선호 메타데이터에 의한 결과를 분석하여 가중치를 변경하고, 좀 더 개인화된 추천 영화를 제공한다.

본 논문은 제 2장에서 메타데이터와 추천시스템에 대한 관련 연구에 대하여 간략하게 설명하고 제 3장에서 본 논문에서 제안하는 온톨로지 기반의 영화 메타데이터 모델에 대하여 설명하고 제 4장에서 제안하는 추천기법에서의 데이터와 구조를 설명하고 제 5장에서 실험과 결과 대하여, 제 6장에서 결론 및 향후 연구에 대하여 기술한다.

## 2. 관련연구

메타데이터는 시스템의 데이터에서 관련 정보들이 결합하여 의미적으로 하나의 데이터를 설명함으로써 메타데이터는 사용자가 각 데이터의 정보를 검색하여 정보를 얻을 수 있도록 돕고, 메타데이터를 통하여 사용자가 찾고자 하는 데이터를 검색할 수 있도록 돕는다(Mathes, 2004). 추천 서비스는 사용자와 아이템의 메타데이터를 활용하여 추천 기법을 적용하고 해당 사용자의 선호도나 사용자와 아이템의 기본 정보를 분석하여 사용자에게 아이템을 추천한다. 본 연구에서는 사용자에게 영화를 추천하기 위하여 영화 메타데이터를 기반으로 영화 메타데이터 온톨로지를 구축하고 영화 메타데이터간의 유사성을 생성하여 영화 추천에 사용한다.

### 2.1 콘텐츠 메타데이터

메타데이터는 데이터를 설명하는 데이터로써 데이터의 검색, 보존, 관리 등의 목적으로 사용한다(Duval et al., 2002). 특히 검색의 경우 특정 도메인의 용어간 연관성을 통하여 추론이 가능한 온톨로지에 적용함으로써 시맨틱한 검색이 가능하도록 발전되고 있다(Vallet et al., 2005). 더블린 코어(Dublin core)는 검색 및 처리가 용이하도록 메타데이터요소를 표준화하여 정의한다(Dublin Core Metadata Initiative, 1999).

콘텐츠 메타데이터는 해당 콘텐츠를 설명하고 콘텐츠의 특징을 잘 나타내며 잘 정의된 콘텐츠 메타데이터는 콘텐츠의 관리를 용이하게 한다(Stamou et al., 2006). 영화 메타데이터는 제목, 장르, 제작국가, 출연자, 감독, 작가, 등급, 평점, 줄거리, 배역 등이 있으며 특히 장르는 영화를 선정할 때 줄거리와 배우와 함께 소비자의 주요 고려사항으로 나타나 있다(Korean Film Council, 2011). 또한 장르는 영화의 검색과 추천을 위한 메타데이터로 활용되고 있다(Lee et al., 2007). 본 논문에서는 영화의 추천에서 사용자의 선호도가 잘 반영되는 장르, 키워드, 줄거리, 배우, 감독, 작가와 영화가 가지는 기본 메타데이터인 제목, 제작국가, 등급, 평점 배역 등의 10가지 영화 메타데이터를 선정하였다. 본 논문에서는 10가지 영화 메타데이터를 기반으로 영화 메타데이터를 표현하고 영화의 추천을 위한 연관성을 도출하며, 이를 표현하기 위해 영화 메타데이터 온톨로지를 만들었다. 제안하는 시스템에서는 영화 추천을 위해 장르뿐만 아니라 줄거리, 배우 등의 메타데이터를 추가함으로써 사용자의 영화 선정 시 고려요인을 반영하여 좀 더 사용자에게 개인화된 결과를 제공하였다.

## 2.2 콘텐츠 추천 시스템

추천 시스템들은 접근하는 방법에 따라 협업적 여과(Collaborative Filtering)와 내용기반 여과(Contents Based Filtering)로 나눌 수 있다. 내용기반 여과는 사용자의 과거 아이템 소비 기록을 기반으로 과거 소비했던 아이템과 유사한 아이템을 추천해 주는 기법이고 협업적 여과는 과거 소비했던 아이템들과 성별, 나이, 직업 같은 사용자의 정보를 기반으로 유사 사용자를 분류하거나 사용자 그룹을 생성하여 과거에 소비했던 아이템을 분석하여 추천해 주는 기법이다. 또한 이 두 기법을 결합한 융합(Hybrid) 기법이 있다(Adomavicius et al., 2005). 현재 상용화 되고 있는 추천 시스템은 내용기반 여과 기법과 협업적 여과 기법을 기반으로 사용자에게 아이템을 추천하며 추천 시스템에서 아이템의 메타데이터들은 사용자가 원하는 아이템을 검색하고 추천될 수 있도록 도와준다.

영화나 TV 프로그램 등의 영상 콘텐츠의 추천 기법들은 사용자의 과거 시청이력을 분석하여 사용자들을 분류하고 비슷한 장르 선호도를 가지는 사용자들의 시청이력을 기반으로 영상 콘텐츠를 추천해 주거나, 영상 콘텐츠의 메타데이터를 분석하여 서로 관련된 여러 개의 유사 영상 콘텐츠 리스트를 분류하여 그룹을 생성하고 사용자의 이력과 사용자 그룹의 정보를 활용하여 사용자의 선호를 분석하고 이를 기반으로 영상 콘텐츠를 추천하여주는 방법이 있다(Kim et al., 2011). 또한, 사용자와 영상 콘텐츠의 메타데이터를 분석하여 사용자와 영상 콘텐츠의 유사도를 도출함으로써 영상 콘텐츠를 추천해 주는 방법이 연구되었다(Tsunoda et al., 2008). 이러한 추천 기법들을 기반으로 사용자들의 소비 패턴을 분석하고 그들의 성별, 나이 같은 기본정보를 기반으로 사

용자들을 분류하고, 분류된 사용자들의 선호도를 분석하여 아이템들을 추천해 주는 기법이 연구되고 있으며 실제 서비스되고 있다(Linden et al., 2003).

본 논문에서는 사용자의 성별, 나이 같은 사용자의 인구 통계학적(Demographic) 정보를 기반으로 사용자 그룹을 나누고 사용자의 그룹별 장르 선호도 기반의 영화를 추천하고 사용자 개인의 선호도를 입력 받아 영화를 검색한다. 또한, 검색된 영화와 유사한 영화를 도출하여 사용자에게 영화를 추천하는 기법을 제안한다.

## 3. 영화 추천을 위한 온톨로지 기반 영화 모델

본 논문에서는 영화의 연관성을 도출하기 위해 영화 메타데이터간의 연관성을 표현하고, 영화추천을 위한 온톨로지 구조의 영화 메타데이터를 표현하는 지식 베이스를 구축한다. 영화 메타데이터 온톨로지는 영화들의 메타데이터와 메타데이터간의 연관성을 표현하기 위한 영화 메타데이터 모델과 사용자의 선호도와 연령, 성별 등의 기본적인 정보를 포함하는 사용자 모델로 구성된다.

온톨로지 구조를 가지는 영화 메타데이터 지식베이스의 구축을 위해 온톨로지 개발 101방법론을 적용한다(Noy et al., 2011). 온톨로지 개발 101방법론은 Stanford 대학 Knowledge Systems 연구실 팀에서 제안한 방법론으로 <Table 1>과 같이 총 7가지 단계로 구성된다. 본 연구에서는 온톨로지를 구축하기 위해 Protégé(Drummond et al., 2005) 온톨로지 편집 툴을 사용하고 영화의 메타데이터의 입력을 위해 Protégé API를 사용하여 자바 프로그램을 구현하였다.

<Table 1> Step of Ontology Development 101

Step	Contents
1	Determine the domain and scope
2	Consider reusing existing ontology
3	Enumerate important terms
4	Define concepts and their hierarchy
5	Define the attributes and relations
6	Define cardinality
7	Create individual

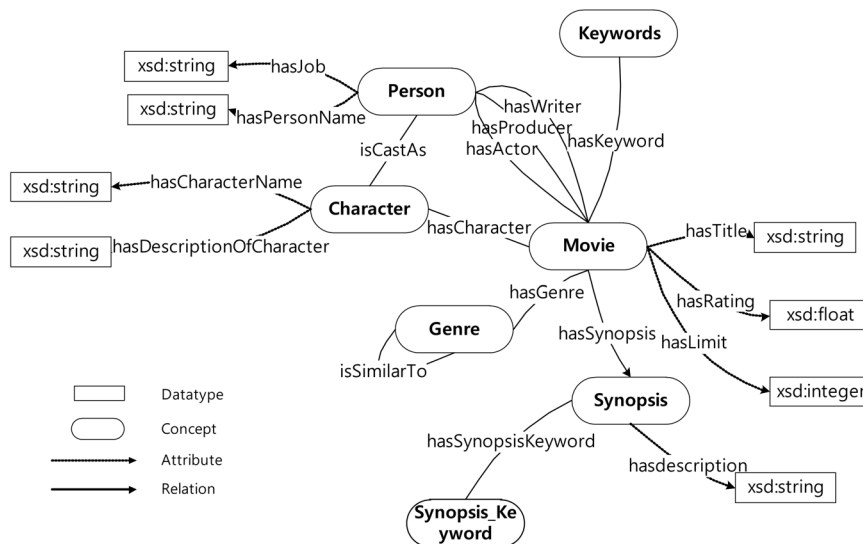
### 3.1 영화 메타데이터 모델

영화 메타데이터 온톨로지 모델은 메타데이터간 연관성을 포함하여 영화간의 유사성을 나타내고 사용자의 선호도를 기반으로 영화 콘텐츠를 검색하고 추천하기 위해 사용한다. 본 논문에서는 영화 메타데이터 온톨로지를 구축하기 위해 영화 콘텐츠의 특징을 잘 나타내는 속성들을 선택하였다. Netflix (<http://www.netflix.com/>), HULU(<http://www.hulu.com/>), Daum 영화(<http://movie.daum.net>)같은 국내외 영화 정보를 제공하는 서비스 업체/포탈에서

영화나 TV 콘텐츠들은 제목, 장르, 출연자(배역), 키워드, 줄거리 같은 메타데이터들을 기반으로 영상 콘텐츠를 분류하고 설명한다. <Figure 1>은 본 논문에서 사용하는 메타데이터들의 개념(Concept)과 속성(Attribute)을 정의하고 개념간의 관계(Relation)를 표현한 온톨로지 모델이다.

<Figure 1>에서 관계는 검정색 화살표로 표현되고 다른 클래스들과의 연관성을 표현한다. 객체의 속성은 점선 화살표로 표현되고 각 클래스들이 가지는 속성들을 표현한다. 영화 메타데이터 모델을 통하여 장르, 배우, 줄거리 같은 영화 메타데이터와 이와 연결되어 있는 관계와 속성을 통하여 영화를 검색할 수 있다.

본 논문에서는 영화 메타데이터 온톨로지를 구현하기 위한 온톨로지 언어로 Web Ontology Language(OWL)를(McGuinness et al., 2004; W3C OWL Working Group, 2009) 사용한다. 각 영화의 메타데이터 들은 영화 메타데이터 모델에서 개념, 속성, 관계로 표현된다.



<Figure 1> Movie Ontology Model

<Table 2> Concept of Movie Metadata

Metadata	Ontology Concept	Description
ID	<i>Contents</i>	Representative Concept of Movie
Staff of Movie	<i>Person</i>	Actors, Directors, Writers
Character	<i>Character</i>	Character of Casted Actor/Actress in Movie
Genre	<i>Genre</i>	Genre of Movie
Keywords	<i>Keyword</i>	Representative Keywords of Movie
Synopsis	<i>Synopsis</i>	Synopsis of Movie
Keywords of synopsis	<i>Synopsis_Keyword</i>	Keywords of KMDB

<Table 3> Relation of Concepts

Subject	Object	Relation
<i>Contents</i>	<i>Person</i>	<i>hasActor</i>
		<i>hasWriter</i>
		<i>hasProducer</i>
	<i>Character</i>	<i>hasCharacter</i>
	<i>Genre</i>	<i>hasGenre</i>
<i>Person</i>	<i>Character</i>	<i>isCastAs</i>
	<i>Genre</i>	<i>isSimilarTo</i>

<Table 4> Attribute of Concepts for Contents Concept

Metadata	Attribute	Description
Title	<i>hasTitle</i>	Title of contents
Limit	<i>hasLimit</i>	Limit of contents
Rating	<i>hasRating</i>	Rating of contents(0.0~10.0)
Country	<i>hasCountry</i>	Country of origin

<Table 2>는 영화 메타데이터 온톨로지를 구축하기 위한 온톨로지의 객체의 개념을 설명한다. 각 객체의 개념들은 관련된 다른 객체의 개념들과 연결은 관계로 표현되며, 영화 메타데이터는 각 객체의 개념마다 연결된 관계를 기반으로 영화를 검색하고 추천할 수 있게 한다. <Table 3>은 객체의 개념으로 표현된 영화 메타데이터들과 각 메타데이터들을 연결하는 관계이다.

Contents개념과 연결되는 다른 객체의 개념들은 영화의 주요 메타데이터들이며 Contents개념과의 연결은 <Table 3>과 같이 관계로 연결된다. Con-

tent스개념과 Person개념 간 연결은 영화와 관련된 배우, 감독, 작가에 따라 연결되는 관계가 다르며 Person개념에서 배우는 해당 영화의 맡은 배역과 isCastAs관계로 연결된다. 또한 Genre개념은 유사장르를 표현하기 위하여 isSimilarTo 관계로 연결된다.

영화의 다른 메타데이터들은 속성으로 표현된다. 속성으로 표현되는 영화 메타데이터는 <Table 4>와 같다.

등급은 사용자의 기본정보에서 연령을 고려하여 영화를 검색하기 위해 사용된다. 평점은 해당 영화를 본 사용자들의 영화 대한 평가를 수치로 표현한 메타데이터로써 추천될 영화의 순위를 결정할 때 고려된다. Contents개념과 연결된 다른 객체의 개념들은 각 객체의 개념들의 정보를 가지는 속성을 가지고 있다. 각 개념이 가지는 속성은 <Table 5>와 같다.

<Table 5> Attribute of Concepts for Other Concepts

Concepts	Attribute	Description
<i>Person</i>	<i>hasPersonName</i>	The name of Actors/Actresses, Directors, Writers
<i>Character</i>	<i>hasCharacterName</i>	The name of Character in Movie
	<i>hasDescriptionOf Character</i>	The description of Character
<i>Synopsis</i>	<i>hasDescription</i>	The Synopsis of Movie

*Person* 개념은 영화에 참여한 사람들의 이름을 속성으로 가지고 있다. 사용자는 *Person* 개념의 속성을 통하여 선호하는 배우, 감독, 작가를 검색하고 이와 연결되어 있는 *Contents* 개념 속성을 통하여 영화를 검색하고 추천 받을 수 있다. *Character* 개념과 연결되는 *hasCharacterName* 속성은 영화의 배역에 대한 설명으로 사용자는 영화에 등장하는 배역에 대한 검색을 통하여 영화를 검색 할 수 있으며 사용자가 선호하는 배역을 통하여 영화를 추천 받을 수 있다.

본 논문에서는 영화 메타데이터 간의 연관성을 도출하기 위해 확률적 방법을 적용한다. 여러 개의 장르를 가지는 영화를 사용하여 조건부확률을 적용한 장르간의 유사성을 도출한다. 장르기반의 유사도 도출을 위한 조건부확률은 식 (1)과 같다.

$$P(C_x|C_y) \quad (1)$$

콘텐츠 모델에서 모든 영화를 대상으로  $C_x$ 는 장르  $x$ 를 가지는 영화를 나타내고  $C_y$ 는 장르  $y$ 를 가지는 영화의 수를 나타낸다. 식 (1)은 콘텐츠 모델에서 축소된 장르  $x$ 에 관한 장르  $y$ 의 상대 확률이다.

<Table 6> Movie Genre

Movie Genre			
Sci-Fi	Family	Horror	Documentary
Drama	Romance	Martial art	Musical
Mystery	Crime	Western	Adult
Thriller	Historical	Animation	Action
Adventure	War	Comedy	Fantasy

본 논문에서는 영화 추천을 위해 20개의 장르를 사용한다. 영화 장르에 대한 각각의 상대 확률을 구하여 장르간의 유사도 값을 계산하였다. 계산된 값을 기반으로 장르  $x$ 와 장르  $y$ 를 유사 장르의 관계로

연결한다. 식 (2)는 각 장르를 연결하기 위한 식이다.

$$R(G_x|G_y) = \begin{cases} 0, & P(C_x|C_y) < 0.2 \\ 1, & P(C_x|C_y) > 0.2 \end{cases} \quad (2)$$

각 장르에 대하여 계산된 조건부 확률 값의 최대값은 0.467이고 0을 제외한 최소값은 0.001이 나왔다. 각 장르별 최대값들의 평균은 0.212가 계산되었다. 계산된 최대값에 따라서 나온 값에 의해 0.212보다 큰 장르들에 대하여 유사 장르로 연결했다. 유사 장르로 계산된 각 장르는 *isSimilarTo* 관계로 연결되며 유사 장르를 통하여 유사 영화 검색과 추천에 사용된다.

영화의 줄거리에 따른 유사도 도출을 위해 Korea Movie Database(KMDB, <http://www.kmdb.or.kr/>)의 9천여 개의 키워드를 기반으로 영화 줄거리의 키워드 사상(Mapping)을 통하여 영화와 키워드를 연결하고 줄거리와 연결된 키워드를 기반으로 영화가 검색되고 추천된다.

### 3.2 사용자 모델

사용자 모델은 사용자에게 영화의 개인화된 추천을 위하여 콘텐츠 모델과 연결하여 하나의 온톨로지 모델로 사용한다. 사용자는 연령과 성별에 따라 사용자 그룹이 결정된다. 각 사용자 그룹은 2011년에 영화진흥위원회에서 조사한 영화 소비자조사를 기반으로 그룹별 선호 장르를 선택하였다(Korean Film Council, 2011). 사용자 그룹은 연령별로 6가지의 그룹과(10대, 20대 초반, 20대 후반, 30대 초반, 30대 후반, 40대, 50대 이상) 성별에 따라 2가지 그룹으로 분류된다.

본 논문에서는 영화의 추천을 위해 사용자가 영화를 선택할 때 고려하는 주요 메타데이터로 줄거리, 장르, 배우, 키워드를 선택하였다. 사용자의 선

호도와 콘텐츠 모델과의 연결을 위해 <Table 7>과 같이 객체의 개념과 관계를 정의하였다.

<Table 7> Relation of Concept for User Model

Subject	Object	Relation
User	Contents Model	prefer_Genre prefer_Actor prefer_Keywords
	Group	belongTo
	Gender	hasGender

사용자 모델에서는 사용자의 아이디를 표현하는 *hasID*와 나이를 표현하는 *hasAge*속성이 있다. *User*는 *hasID*와 *hasAge* 속성과 <Table 7>에서 정의된 관계를 기반으로 표현되며 각 객체의 개념들의 연결은 <Figure 2>와 같다.

<Figure 2>는 사용자 모델에서의 객체의 개념과 관계, 속성들을 보여준다. 사용자 모델에서 *User*는 선호 메타데이터에 따라 선호 장르, 배우, 줄거리, 키워드의 관계를 가지고 각 관계는 영화 모델의 장르, 배우, 줄거리, 키워드와 연결된다. 선호 장르의 연결에서 사용자의 선호 장르는 유사 장르와도 연결된다. *UserGroup*개념은 앞서 정의한 연령별 6가

지와 성별 2가지 그룹을 가지고 있으며, 각 그룹별 선호 장르가 영화 모델의 장르와 연결된다.

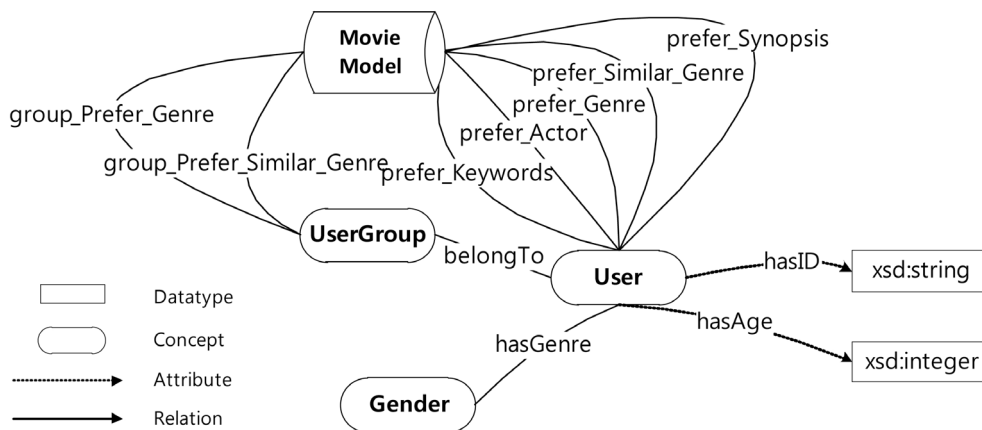
## 4. 데이터 입력과 추천시스템 구조

### 4.1 데이터 입력

영화 메타데이터의 정보 입력을 위해 국내 포털 사이트(<http://movie.daum.net>)에서 제공하는 영화와 각 영화의 메타데이터를 사용한다. 본 논문에서 사용하는 영화의 수와 각 개념에서의 개체(Individual)의 수는 <Table 8>과 같다.

<Table 8> The Number of Individual

Concept	The Number of Individual	Description
Movie	14,374	The Number of Movie
Character	52,403	The Number of Character
Person	28,026	The Number of Staff
Synopsis_Keyword	9,472	The Number of Keywords (Consider with the Synopsis)
Genre	20	The Number of Genre
Keywords	188	The Number of Keywords



<Figure 2> User Model



## 4.2 영화 추천 시스템 구조

본 논문에서는 영화 추천을 위하여 영화 메타데이터 온톨로지를 기반으로 사용자의 연령, 성별 등을 이용하여 사용자의 그룹을 결정하고 그룹의 장르 선호도에 따라 영화를 검색한다. 그리고 사용자의 선호도를 입력 받아 영화를 검색하고 사용자에게 추천한다.

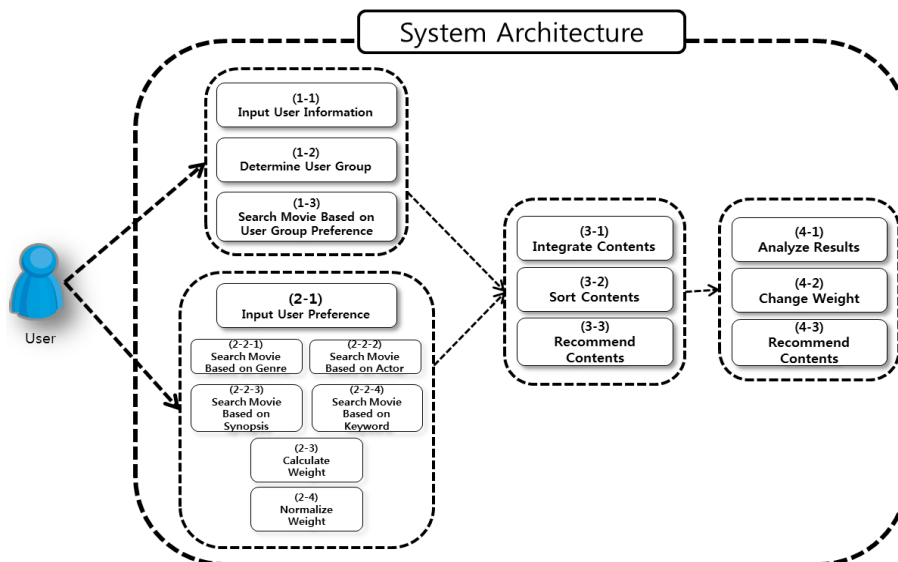
<Figure 3>은 영화의 추천을 위한 시스템 구조를 보여준다. 제안하는 영화추천 시스템은 4가지 컴포넌트로 정의 된다. 1) 입력 받은 사용자의 연령, 성별을 기반으로 사용자의 그룹을 결정하고 그룹의 장르 선호도에 따른 후보 영화의 검색, 2) 사용자가 선호하는 장르, 배우, 줄거리, 키워드 같은 영화의 메타데이터를 입력 받아 후보 영화를 검색 3) 검색된 후보 영화에 가중치를 부여하여 영화를 정렬하고 사용자에게 영화를 추천, 4) 사용자가 선택한 메타데이터와 추천된 영화를 분석하여 메타데이터의 가중치를 변경하고 영화를 추천하는 컴포넌트로 구

성된다.

후보 영화 검색을 위해 정의된 영화 메타데이터 온톨로지를 기반으로 사용자 그룹의 선호 장르와 사용자가 입력한 선호 장르, 배우, 줄거리, 키워드를 기반으로 하여 해당 메타데이터와 연결된 영화 들을 검색한다. 검색된 후보 영화들은 메타데이터에 따라 정의된 가중치를 계산하여 각각의 영화의 가중치를 계산하고, 계산된 가중치를 기반으로 영화를 정렬하고 추천한다.

## 5. 영화 추천을 위한 메타데이터 활용

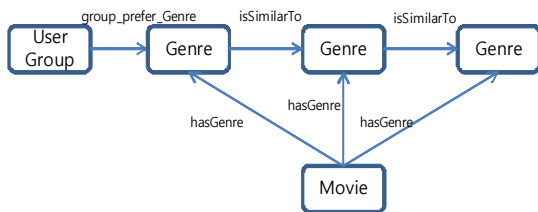
본 논문에서는 영화의 추천을 하기 위하여 성별, 연령 같은 인구 통계학적(Demographic) 정보를 활용한 영화의 추천과 사용자가 선호하는 장르, 배우, 줄거리, 키워드 같은 영화 메타데이터를 기반으로 영화를 검색하고 추천한다. 검색된 후보 영화는 가중치 계산을 통하여 가중치 값에 따라 사용자에게 영화를 추천해 준다.



<Figure 3> Movie Recommendation System Architecture

### 5.1 사용자 정보기반 영화 검색

사용자 정보기반 영화 검색은 사용자의 기본 정보를 기반으로 사용자의 그룹을 결정하고 그룹의 장르 선호도와 *isSimilarTo* 관계로 연결된 유사장르에 따라서 영화를 검색한다. 사용자의 정보가 입력되면 정의된 Semantic Web Rule Language(SWRL)(Horrocks et al., 2004) 규칙에 따라 그룹이 결정되고 사용자가 속한 그룹의 선호 장르와 선호 장르의 유사장르를 기반으로 쿼리를 생성하여 추천될 후보 영화를 검색한다. 사용자 그룹은 연령별로 8가지, 성별로 2가지로 구분 된다. 각 사용자 그룹은 ‘영화진흥위원회’에서 조사한 자료(Korean Film Council, 2011)를 기반으로 그룹별 선호 장르를 설정하였다. 사용자 정보 기반의 영화 검색을 위한 쿼리는 <Figure 4>와 같이 개념들의 연결된 관계를 통하여 영화를 검색한다. *UserGroup* 개념의 *group\_Prefer\_Genre* 관계를 통하여 정의된 선호 장르를 검색하고 선호장르와 유사 장르로써 *isSimilarTo* 관계로 연결된 다른 장르를 검색하여 각 장르와 연결된 영화를 검색한다. 유사 장르의 관계는 총 두 번의 링크까지 허용한다. 예를 들어 사용자 그룹이 선호하는 장르가 ‘전쟁’이라면 ‘전쟁’과 유사 장르로 연결된 장르 ‘액션’과 장르 ‘액션’과 연결된 ‘스릴러’까지 도출되고 각 장르와 연결된 영화가 검색 된다.



<Figure 4> Searching the Movie for each Group

사용자 그룹의 선호 장르와 선호 장르와 유사한 두 개의 장르를 기반으로 검색된 영화는 영화 추천을 위한 후보 영화 리스트가 된다. 각 후보 영화는 검색된 조건에 따라 가중치를 다르게 가지며 검색된 영화의 추천을 위한 가중치는 영화 메타데이터 온톨로지에 입력된 평점(Rating)값과 함께 계산되어 영화의 가중치를 결정한다. 사용자 정보기반 영화 검색에서 검색된 후보 영화 리스트들의 가중치 계산은 식 (3)과 같다.

$$W(C_{xy}) = \sum_{x=genre_i} W(G_x) + C_{xy}Rating \quad (3)$$

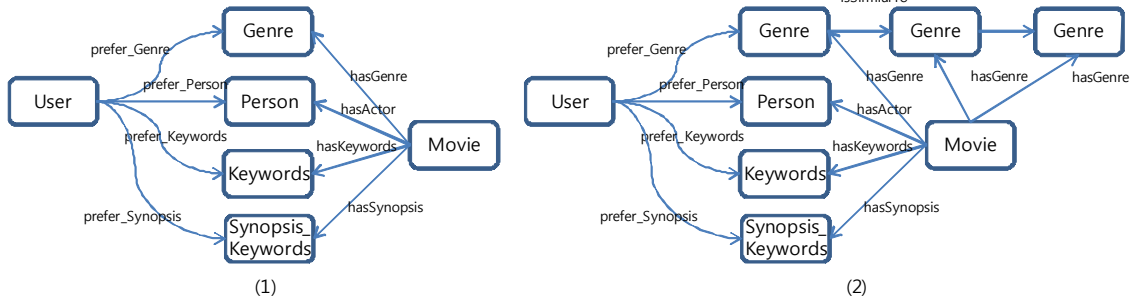
영화  $C_{xy}$ 는 장르  $x$ 로 검색된 영화  $y$ 이고,  $C_{xy}$ -*Rating*은 장르  $x$ 로 검색된 영화  $y$ 의 평점이다. 장르  $G_x$ 의 가중치는 식 (4)에 의해 정의된다.

$$W(G_x) = \begin{cases} 0.8 & G_x = \text{그룹의 선호 장르} \\ 0.5 & G_x = \text{그룹의 선호 장르와 유사한 장르} \\ 0.3 & G_x = \text{유사 장르와 유사한 장르} \end{cases} \quad (4)$$

그룹의 선호 장르와 선호 장르와 유사한 장르를 기반으로 검색된 후보 영화 리스트는 식 (3)과 식 (4)에 의하여 가중치 값을 가지며, 가중치 값은 후보 영화가 사용자에게 추천될 때 영화의 순위를 결정한다.

### 5.2 영화 선호 메타데이터 기반 영화 검색

사용자가 영화를 선택할 때 주요 고려 사항으로는 장르, 배우, 줄거리, 키워드가 있다. 본 논문에서는 사용자가 영화를 선택할 때 고려하는 메타데이터를 기반으로 사용자가 원하는 영화를 검색한다. 사용자는 주요 메타데이터를 입력하고 입력한 메타데이터를 기반으로 영화 메타데이터



<Figure 5> Searching the Movie Based on user Preference (1) Relation of Concept with no Similar Genre, (2) Relation of Concept with Similar Genre

온톨로지에서 영화를 검색한다. <Figure 5>는 영화 모델에서 영화의 검색을 위한 *User* 개념과 영화 모델에서의 주요 개념과의 관계에 대한 연결을 보여준다. <Figure 5>에서 (1)은 사용자가 입력한 선호 영화 메타데이터를 기반으로 연결된 영화를 검색할 때의 연결을 보여준다. (2)는 영화 메타데이터 온톨로지 내에서 사용자가 입력한 선호 영화 메타데이터의 장르 중 유사장르로 연결된 장르들을 확정하여 영화를 검색할 때의 연결을 보여준다.

검색된 영화 리스트는 사용자에게 추천되기 위한 후보 영화다. 각각의 후보 영화 리스트는 추천을 위한 가중치 값을 갖는다. 사용자 선호도 기반으로 검색된 영화의 가중치 값의 계산은 식 (5)와 같이 정의된다.

$$W(C_k) = nor\left(\sum_{i=1}^n W(G_i)\right) + nor\left(\sum_{i=1}^n W(A_i)\right) + nor\left(\sum_{i=1}^n W(SK_i)\right) + nor\left(\sum_{i=1}^n W(K_i)\right) \quad (5)$$

$n$ 은 사용자가 선호하는 메타데이터로 선택한 장르, 배우, 줄거리, 키워드의 수이다. 사용자의 선호도에 의해 검색된 영화  $C_k$ 의 가중치 값은 영화  $G_k$ 와

연결되어 있는 메타데이터에서 사용자가 선호하는 영화 메타데이터들의 가중치 합으로 계산한다.  $C_k$ 의 가중치 계산을 위해 사용된 메타데이터들의 가중치는 해당 영화에서의 동일한 범위를 가지기 위해 1~2로 정규화 된다. 각각의 메타데이터의 정규화는 식 (6)과 같다.

$$nor(C_{km}) = \begin{cases} \frac{value(M_m) - \min(M_m)}{\max(M_m) - \min(M_m)} + 1 & \max(M_m) \neq \min(M_m) \\ 1.5 & \max(M_m) = \min(M_m) \end{cases} \quad (6)$$

식 (6)에서  $C_{km}$ 은 메타데이터  $m$ 으로 검색된 영화  $G_k$ 다.  $m$ 은 사용자가 영화를 선택할 때 고려하는 주요 메타데이터로  $m = \{\text{장르, 배우, 줄거리, 키워드}\}$ 이다.  $M_m$ 은 메타데이터  $m$ 으로 검색된 영화 리스트  $C$ 에서 각각의 메타데이터의 가중치 값이다.

식 (5)에 의해 계산된 가중치를 기반으로 추천된 영화는 사용자가 선택한 메타데이터에 의한 추천결과와 함께 분석되어 각 메타데이터의 가중치를 변경한다. 각 가중치의 변경은 <Table 9>와 같으며 메타데이터의 가중치 적용은 식 (7)과 같다. 각 메타데이터에 의해 추천된 결과의 기여도를 기반으로 각 메타데이터에 가중치를 곱하여 새로운 가중치를 가지게 된다.

$$W(C_k) = nor\left(\sum_{i=1}^n W(G_i)\right) \times CW_G + nor\left(\sum_{i=1}^n W(a_i)\right) \quad (7)$$

$$\times CW_A + nor\left(\sum_{i=1}^n W(SK_i)\right) \times CW_{SK}$$

$$+ nor\left(\sum_{i=1}^n W(K_i)\right) \times CW_K$$

각 메타데이터의 가중치에 곱해지는 기여도  $CW$ 는 사용자가 선택한 선호 메타데이터에 의해 검색되어 추천된 영화와 결과에 의해 분석된 결과에 의해 추천 결과에 영향을 미치는 기여도 순으로 각각의 메타데이터의 기여도 순에 따라  $CW$ 는 2, 1.6, 1.3, 1 가진다.

본 논문에서 제안하는 영화 추천은 영화 메타데이터 온톨로지에 정의된 유사 메타데이터를 통하여 사용자가 선택한 메타데이터뿐만 아니라 입력하지 않은 메타데이터에 의한 검색이 가능하다. 검색된 영화 리스트는 사용자의 선호도에 따라 영화의 순위를 가지게 되고 사용자는 순위가 정해진 영화에 따라 영화를 추천 받는다.

## 6. 실험

### 6.1 실험데이터 및 평가 방법

본 논문에서는 제안하는 방법의 실험을 위하여 국내 포털사이트(Daum 영화)에서 제공하는 영화 메타데이터를 사용하였다. 평점 7.0 이상을 가지는 영화 7,418개를 사용하며 2011년 ‘영화진흥위원회’에서 조사한 결과에서 영화관람의 비중이 가장 높은 임의의 20대 남/녀(Korean Film Council, 2011) 10명씩 20명을 대상으로 실험하였다. 실험을 위해 사용한 영화의 주요 메타데이터의 사이즈는 <Table 9>와 같다.

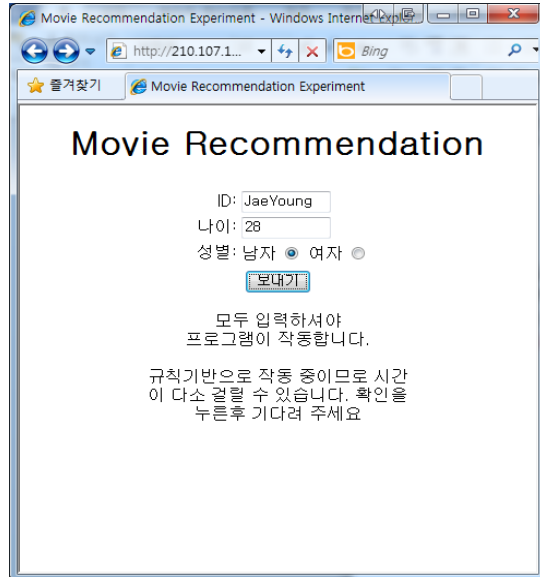
사용자의 정보와 선호도를 기반으로 Top-5, Top-10, Top-20의 영화를 추천하고 추천된 영화 목록에서 만족하는 영화를 입력 받았다. 각 메타데이터에

대한 선호도의 선택은 각 메타데이터에 의해 검색될 수 있는 영화에 의해 추천될 수 있는 영향을 줄이기 위해 5개 이하로 제한하였다.

<Table 9> The Number of Individual for Experiment

Concept	The Number of Individual
Movie	7,418
Actor	12,771
Synopsis	7,418
Synopsis_Keyword	9,472
Keywords	188
Genre	20

<Figure 6>~<Figure 8>은 사용자를 대상으로 한 실험을 위해 구현된 웹 프로그램의 구현 화면이다. <Figure 6>은 사용자의 ID와 나이, 성별을 입력하는 화면이다. <Figure 7>은 사용자가 선호하는 메타데이터를 입력하는 화면이며, 마지막으로 <Figure 8>은 입력에 대한 영화 추천 결과 화면이다.



<Figure 6> Demo Capture for Input the User Information

온톨로지 기반 영화 메타데이터간 연관성을 활용한 영화 추천 기법



<Figure 7> Demo Capture for Input the User Preference

동향연 검색결과1				동향연 검색결과2			
제목	장르	감독	평점	제목	장르	감독	평점
1 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	1 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
2 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	2 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
3 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	3 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
4 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	4 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
5 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	5 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
6 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	6 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
7 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	7 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
8 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	8 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
9 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	9 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
10 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	10 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
11 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	11 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
12 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	12 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
13 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	13 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
14 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	14 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
15 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	15 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
16 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	16 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
17 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	17 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
18 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	18 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
19 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	19 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
20 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	20 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
21 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	21 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
22 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	22 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
23 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	23 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
24 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	24 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
25 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	25 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
26 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	26 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
27 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	27 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
28 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	28 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
29 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	29 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5
30 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5	30 1907.1907.1907	액션, 스릴러	제임스 카메론	8.5

<Figure 8> Demo Capture for Recommendation Result

## 6.2 실험결과

제안하는 영화 추천 기법을 평가하기 위하여 각 사용자들은 성별과 연령을 입력하고 선호하는 장르, 배우, 키워드, 줄거리를 입력하고 영화를 추천 받는다.

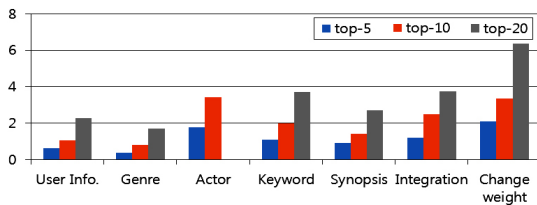
<Table 10>과 <Figure 9>은 영화가 추천되었을 때 성별, 연령에 따른 사용자의 선호 장르기반의 추천과 사용자의 선호 메타데이터기반의 추천, 각 결과들을 합친 후의 영화 추천결과와 가중치를 변경후의 결과를 보여준다. 제안하는 방법에 의해 추천된 영화 중 Top-20에서의 선택된 영화의 평균수는 6.35개로 가장 높게 나왔고 장르에 의해 추천된 영화 중 Top-20에서의 선택된 영화의 평균수는 1.7개로 가장 낮게 나왔다. 또한, 배우의 경우 실험에 참여한 피실험자가 입력한 배우들이 출연한 영화의 수가 15개 이상의 결과가 나오지 않아 Top-20의 결과가 0이 나왔다.

<Table 11>과 <Figure 10>은 추천된 영화의 수

와 사용자가 선택한 영화수의 비율을 보여준다. 실험 결과 제안하는 방법에 의해 추천된 영화의 만족도가 가장 높게 나왔으며, 영화 추천에 가장 많이 사용되는 장르의 경우 사용자에게 충분한 영향을 주지 못하는 것으로 나타났다. 장르를 사용하여 영화를 검색할 경우 각 장르에는 많은 양의 영화를 포함하고 있기 때문에 사용자에게 만족하는 영화를 제공하기가 힘들었다. 이를 보완하기 위해 영화 메타데이터에서 사용자가 영화를 선택할 때 고려하는 다른 특정 메타데이터들로 확장하여 영화를 추천하였을 때 사용자의 만족도를 높이 이끌어 낼 수 있었다. 또한 사용자의 특정 메타데이터에 대한 선호도를 분석하여 가중치를 다르게 주어 추천하였고, 사용자에게 높은 정확도를 줄 수 있었다. 하지만 사용자 개인의 선호 메타데이터에 대한 가중치를 동일하게 준 결과 사용자에 따라 영화를 선택할 때 각각의 메타데이터에서 더 선호하는 메타데이터에 대한 정확한 선호 메타데이터가 반영되지 않았다.

<Table 10> Average Number of Preferential Movie for Recommended Movie Based on Metadata

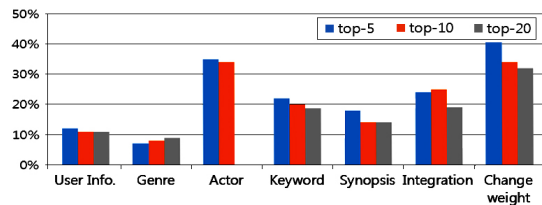
	Top-5	Top-10	Top-20
User Info.	0.6	1.05	2.25
Genre	0.35	0.8	1.7
Actor	1.75	3.4	-
Keyword	1.1	2.0	3.7
Synopsis	0.9	1.4	2.7
Integration	1.2	2.5	3.75
Change weight	2.1	3.35	6.35



<Figure 9> Average Number of Preferential Movie for Recommended Movie Based on Metadata

<Table 11> Recommendation Accuracy for Each Metadata(%)

	Top-5	Top-10	Top-20
User Info.	12%	11%	11%
Genre	7%	8%	9%
Actor	35%	34%	-
Keyword	22%	20%	19%
Synopsis	18%	14%	14%
Integration	24%	25%	19%
Change weight	42%	34%	32%



<Figure 10> Recommendation Accuracy for Each Metadata

## 7. 결론 및 향후 연구

본 논문에서는 영화 메타데이터의 온톨로지를 만들고 메타데이터간의 연관성을 도출하여 영화 메타데이터 온톨로지에 표현함으로써 사용자가 원하는 영화를 검색하고 추천해주는 시스템을 제안하였다. 본 논문에서는 영화의 메타데이터를 기반으로 영화간의 연관성을 도출하고 사용자의 선호도를 입력받아 사용자의 개인화된 정보를 활용하여 영화를 추천하는 기법을 제안하였다. 본 논문에서 제안하는 기법을 평가하기 위해 7,418개의 영화를 사용하여 영화 메타데이터 온톨로지를 구축하고 실험하였다. 사용자는 사용자의 나이와 성별을 통하여 사용자 그룹이 결정되고 사용자 그룹에 따라 선호하는 장르기반의 영화를 검색하고 추천받게 되고, 사용자의 장르, 배우, 키워드, 줄거리를 입력받아 해당 메타데이터와 연결된 영화를 검색하고 추천받게 된다. 각 메타데이터에 의해 검색된 결과에서 각각 같은 가중치를 주었을 때 배우의 경우 통합된 결과보다 좋은 결과를 가져왔다. 따라서, 본 논문에서는 각 결과를 분석을 통해 각 메타데이터에 대한 사용자의 선호도를 반영하였다. 실험 결과 제안하는 기법에서 영화를 추천하였을 때 장르 등의 메타데이터를 기반으로 하여 추천하였을 때 보다 좋은 결과가 나왔으며, 추천된 영화 중 만족하는 영화의 평균 수는 Top-5, Top-10, Top-20에서 각각 2.1개, 3.35개, 6.35개가 나왔다. 이는 각 메타데이터를 기반으로 추천된 결과보다 평균 18% 좋은 결과를 가져왔다. 또한 다른 영화 추천시스템이나 영상 콘텐츠 추천 시스템에서 사용하는 메타데이터인 장르의 경우 다른 메타데이터에 비하여 낮은 만족도를 보여 주었다.

본 연구의 영화 추천 기법은 사용자의 정보를 활용하여 사용자의 그룹을 정의하고, 사용자가 속한

그룹이 선호하는 영화 메타데이터와 사용자 개인의 선호도를 기반으로 개선된 영화의 추천기법을 제안했다. 또한, 영화 메타데이터간의 연관성 도출을 통하여 영화간의 유사성을 도출하고 이를 기반으로 사용자의 기본적인 정보를 활용한 추천뿐만 아니라 사용자가 예상하지 못한 영화의 추천이 가능하다.

향후 연구로 사용자의 영화 시청 이력 분석을 통하여 사용자의 특정 메타데이터에 대한 선호도를 분석하여 각 메타데이터에 대한 가중치를 다르게 적용함으로써 사용자의 선호 메타데이터를 정확하게 반영할 수 있고, 사용자가 시스템을 통하여 추천 영화를 제공받기 위한 노력을 최소화 하여 자동화된 영화 추천 기법을 연구하고자 한다. 또한 본 연구의 실험에서 제한된 실험자의 수로 통계학적 의미의 결과를 가져오지 못하였다. 따라서, 많은 수의 실험을 통하여 제안하는 기법을 실험하고자 한다.

## 참고문헌

- Adomavicius, G. and A. Tuzhilin, "Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions," *Knowledge and Data Engineering*, Vol.17, No.6(2005), 734~749.
- Bechhofer, S., F. Van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, *OWL web ontology language reference*, W3C recommendation 10, 2004. Available at <http://www.w3.org/TR/owl-ref/>(Accessed 15 June 2013).
- Chen, T., W. L. Han, H. D. Wang, Y. X. Zhou, B. Xu, and B. Y. Zang, "Content recommendation system based on private dynamic user profile," *Machine Learning and Cybernetics*, Vol.4(2007), 2112~2118.

- DAUM Movie of Korea homepage. Available at <http://movie.daum.net>(Accessed 20 June 2013).
- Drummond, N., M. Horridge, and H. Knublauch, "Protégé-OWL tutorial," *8th International Protégé Conference*, (2005).
- Dublin Core Metadata Initiative, *Dublin core metadata element set, version 1.1 : Reference description*, 1999. Available at <http://dublin-core.org/documents/2003/08/26/usageguide> (Accessed 17 June 2013).
- Duval, E., W. Hodgins, S. Sutton, and S. L. Weibel, *Metadata principles and practicalities*, *D-lib Magazine*, 2002. <http://www.dlib.org/dlib/april02/weibel/04weibel.html>(Accessed 15 June 2013).
- Gunawardana, A. and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *Journal of Machine Learning Research*, Vol.10(2009), 2935~2962.
- Horrocks, I., P. F. Patel-Schneider, H. Boley, and S. Tabet, *SWRL : A semantic web rule language combining OWL and RuleML*, *W3C Member submission*, 2004. Available at <http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/>(Accessed 28 June 2013).
- HULU of U.S homepage. Available at <http://www.hulu.com>(Accessed 25 June 2013).
- Kim, E., S. Pyo, E. Park, and M. Kim, "An automatic recommendation scheme of TV program contents for (IP) TV personalization," *Broadcasting IEEE Transaction*, Vol.57, No.3 (2011), 674~684.
- Korean Film Council(Korea), *Report for Movie Consumption*, Korean Film Council, 2011.
- Korean Movie Database KMDb. Available at <http://www.kmdb.or.kr>(Accessed 25 June 2013).
- Lee, J. S. and S. D. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering," *Journal of Intelligence and Information Systems*, Vol.13, No.4(2007), 65~78.
- Linden, G., B. Smith, and J. York, "Amazon.com recommendations : Item-to-item collaborative filtering," *Internet Computing*, Vol.7, No.1 (2003), 76~80.
- Mathes, A., "Folksonomies-cooperative classification and communication through shared metadata," *Computer Mediated Communication*, Vol.47, No.10(2004).
- McGuinness, D. L. and F. Van Harmelen, *OWL web ontology language overview*, W3C recommendation 10, 2004. Available at [www.w3.org/TR/owl-features](http://www.w3.org/TR/owl-features)(Accessed 13 June 2013).
- NETFLIX of U.S homepage. Available at <http://www.netflix.com>(Accessed 25 June 2013).
- Noy, N. F. and D. L. McGuinness, *Ontology development 101 : A guide to creating your first ontology*, Stanford Knowledge Systems Laboratory Tech, (2001).
- Stamou, G., J. van Ossenbruggen, J. Z. Pan, G. Schreiber, and J. R. Smith, "Multimedia annotations on the semantic web," *Multimedia IEEE*, Vol.13, No.1(2006), 86~90.
- Su, X. and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in Artificial Intelligence*, (2009), 1~20.
- Tintarev, N. and J. Masthoff, "A survey of explanations in recommender systems," *Proceedings of the IEEE 23rd International Conference on Data Engineering Workshop*, (2007), 801~810.
- Tsunoda, T. and M. Hoshino, "Automatic metadata expansion and indirect collaborative filtering for TV program recommendation system," *Multimedia Tools and Applications*, Vol.36, No.1-2(2008), 37~54.
- Vallet, D., M. Fernández, and P. Castells, "An



- ontology-based information retrieval model,” *The Semantic Web : Research and Applications 2nd European Semantic Web Conference*, (2005), 455~470.
- Weng, S. S. and M. J. Liu, “Personalized product recommendation in e-commerce,” *2004 IEEE International Conference on e-Technology, e-Commerce, and e-Services*, (2004), 413~420.
- W3C Owl Working Group, *OWL 2 web ontology language document overview*, W3C Recommendation, 2009. Available at <http://www.w3.org/TR/owl-overview>(Accessed 13 June 2013).
- Yun, B. D., J. W. Kim, Y. S. Cho, and S. G. Kang, “Personalized Recommendation System for IPTV using Ontology and K-medoids,” *Journal of Intelligence and Information Systems*, Vol.16, No.3(2010), 147~161.

Abstract

## The Ontology Based, the Movie Contents Recommendation Scheme, Using Relations of Movie Metadata

Jaeyoung Kim\* · Seok-Won Lee\*\*

Accessing movie contents has become easier and increased with the advent of smart TV, IPTV and web services that are able to be used to search and watch movies. In this situation, there are increasing search for preference movie contents of users. However, since the amount of provided movie contents is too large, the user needs more effort and time for searching the movie contents. Hence, there are a lot of researches for recommendations of personalized item through analysis and clustering of the user preferences and user profiles. In this study, we propose recommendation system which uses ontology based knowledge base. Our ontology can represent not only relations between metadata of movies but also relations between metadata and profile of user. The relation of each metadata can show similarity between movies. In order to build, the knowledge base our ontology model is considered two aspects which are the movie metadata model and the user model. On the part of build the movie metadata model based on ontology, we decide main metadata that are genre, actor/actress, keywords and synopsis. Those affect that users choose the interested movie. And there are demographic information of user and relation between user and movie metadata in user model. In our model, movie ontology model consists of seven concepts (Movie, Genre, Keywords, Synopsis Keywords, Character, and Person), eight attributes (title, rating, limit, description, character name, character description, person job, person name) and ten relations between concepts. For our knowledge base, we input individual data of 14,374 movies for each concept in contents ontology model. This movie metadata knowledge base is used to search the movie that is related to interesting metadata of user. And it can search the similar movie through relations between concepts.

We also propose the architecture for movie recommendation. The proposed architecture consists

---

\* Computer Science, Ajou University

\*\* Corresponding Author: Seok-Won Lee

Department of Software Convergence Technology, Ajou University,  
206, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-749, Korea  
Tel: +82-31-219-2434, Fax: +82-31-219-1621, E-mail: leesw@ajou.ac.kr

of four components. The first component search candidate movies based the demographic information of the user. In this component, we decide the group of users according to demographic information to recommend the movie for each group and define the rule to decide the group of users. We generate the query that be used to search the candidate movie for recommendation in this component.

The second component search candidate movies based user preference. When users choose the movie, users consider metadata such as genre, actor/actress, synopsis, keywords. Users input their preference and then in this component, system search the movie based on users preferences. The proposed system can search the similar movie through relation between concepts, unlike existing movie recommendation systems. Each metadata of recommended candidate movies have weight that will be used for deciding recommendation order.

The third component the merges results of first component and second component. In this step, we calculate the weight of movies using the weight value of metadata for each movie. Then we sort movies order by the weight value.

The fourth component analyzes result of third component, and then it decides level of the contribution of metadata. And we apply contribution weight to metadata. Finally, we use the result of this step as recommendation for users.

We test the usability of the proposed scheme by using web application. We implement that web application for experimental process by using JSP, Java Script and protégé API. In our experiment, we collect results of 20 men and woman, ranging in age from 20 to 29. And we use 7,418 movies with rating that is not fewer than 7.0. In order to experiment, we provide Top-5, Top-10 and Top-20 recommended movies to user, and then users choose interested movies. The result of experiment is that average number of to choose interested movie are 2.1 in Top-5, 3.35 in Top-10, 6.35 in Top-20. It is better than results that are yielded by for each metadata.

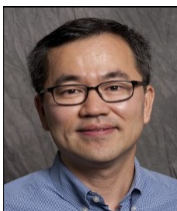
**Key Words** : Recommendation System, Metadata, Ontology, Movie Contents

## 저자 소개



김재영

을지대학교 의료전산학과를 졸업하고 현재 아주대학교 석사과정에 재학 중이다. 주요 관심분야는 데이터마이닝, 지식표현, 기계학습, 추론 등이다.



이석원

현재 아주대학교 소프트웨어 융합학과 교수로 재직 중이다. 동국대학교 컴퓨터공학과를 졸업하고, University of Pittsburgh에서 Computer Science 석사학위를 취득하였고, George Mason University에서 Computer Science 박사학위를 취득하였다. IBM Thomas J. Watson Research Center, Science Application International Corporation에서 연구원으로 활동하였으며 The University of North Carolina at Charlotte Software and Information 조교수, University of Nebraska-Lincoln Computer Science and Engineering 초빙교수, The University of Texas at San Antonio Information systems and Cyber Security 조교수로 활동 하였다. ACM Transactions on Software Engineering Methodology (TOSEM), Requirements Engineering (RE) Journal, IEEE Transactions on Software Engineering, IEEE Transaction SMC 등에서 심사위원으로 활동하였다. 현재 Requirements Engineering Conference의 Program Committee 멤버로 활동 중이다. 연구 관심분야는 기계학습, 인공지능, 의료정보, 요구공학, 소프트웨어공학 등이다.