

Constructing a Knowledge Base for Entertainment by Interlinking Multiple Data Sources

Haklae Kim¹ and Jungyeon Yang²

¹ Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon, 34141, Korea

² Samsung Electronics Co., Ltd., (Maetan dong) 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16677, Korea

Abstract

This paper describes a knowledge base for entertainment domains, including movies, music, and celebrities. We present an ontology model for representing graph-based knowledge, and describe knowledge processing techniques for constructing an entertainment knowledge base, which is comprised of collection and extraction of large-scale data, and knowledge transformation from them. In particular, we address entity resolution and consolidation techniques for integrating particular entities among heterogeneous data sources. We also demonstrate a mobile application using this knowledge base, which allows users to discover relevant entertainment content based on question & answer system.

Keywords: Knowledge base; knowledge graph; linked data; entertainment.

Received on 21 November 2016, accepted on 30 December 2016, published on 01 February 2017

Copyright © 2017 Haklae Kim and Jungyeon Yang, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.1-2-2017.152152

1. Introduction

As various entertainment and multimedia content services on social and mobile environments have become popular, individual service should make distinctive features compared to others. One of fundamental challenges is how they are able to deliver a right content to their users when is needed [1][2]. There have been a huge number of research efforts for this subject. Recently, personalisation and recommendation based on big data analytics of users' behaviours would be one of candidates for providing better features [3]. Simultaneously, we also need to consider characteristics of entertainment content and average usages of users.

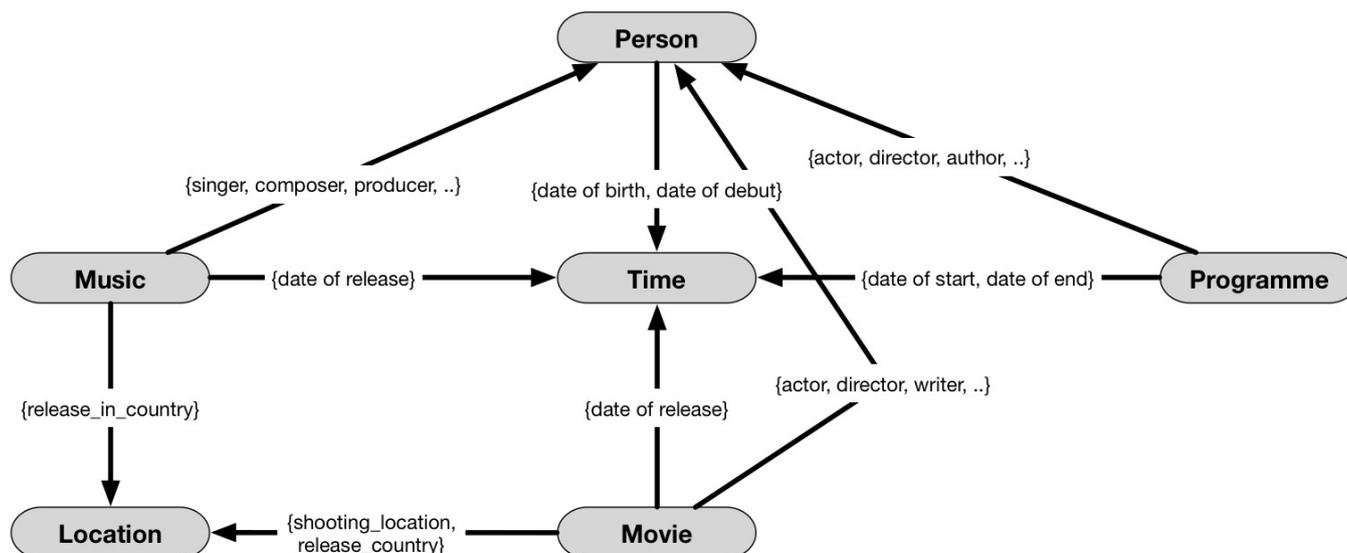
Most of users, in general, tend to discover a set of relevant content across multiple domains such as movies and music. For example, when users watch a movie, they would want to listen to an original soundtrack or to know about actors or actress of that movie. Let us imagine the following query: *a film that is related to a musician, who sings "Just love it."* To obtain a right answer, a system should investigate some relationships among required information such as person, music and movie. Then, the system provides: the answer:

Eminem is the musician, and he appears on a list of films such as How to Make Money Selling Drugs (2013), Something from Nothing: The Art of Rap (2012), Have Gun - Will Travel (2008), 50 Cent: Bulletproof (2005), 8 Mile (2002) and The Wash (2001).

Interlinking content entities among heterogeneous data sources is essential for cross-domain discovery and search. Although there are a huge number of applications and services for content recommendation [4], only few applications allows to discover a set of domain content across different data sources Thus, to enable hybrid content discovery, a knowledge base should have various domain content and they also interconnect among them by a consistent manner.

*Corresponding author. Email: haklaekim@gmail.com

Figure 1 Core classes of the entertainment knowledge model. Each class is connected to others with specific properties.



In this study, we describe a knowledge base for discovering cross-domain content by a single query or request. Most of objects in a knowledge base have some associations among others, and they are interconnected to others at a semantic level [4]. We also introduce knowledge processing techniques for constructing large-scale entertainment knowledge, and demonstrate a prototype system that uses this knowledge base. In particular, this system provides advanced features such as question answering for user interface, contextual recommendation based on semantic search.

Based on the knowledge base, a question answer engine is able to understand users’ complex questions from multiple domains. In summary, we make the following contributions:

- We introduce an integrated knowledge base, which is created by large-scale curated and extracted data from heterogeneous sources.
- We describe a process for constructing a knowledge base using linked data technologies from data collection to knowledge transformation including knowledge interlinking across multiple data sources.
- We demonstrate a mobile application, which provides an effective recommendation using the knowledge base.

In Section 2, we describe related work in more detail before moving on to overview of knowledge processing techniques (Section 3) and a mobile application system as a use case of the entertainment knowledge base (Section 4).

2. Related work

In past few years, knowledge bases with a graph structure have rapidly increasing of both commercial and public services. Freebase is a free, knowledge graph with millions

of global and general information about real world entities, including well-known people, places and things [5][6]. Some of commercial knowledge bases, such as Google Knowledge Graph and Microsoft Bing, was powered in part by Freebase as a base knowledge, because Freebase data was available for commercial and non-commercial use under a Creative Commons Attribution License. Freebase contained data harvested from sources such as Wikipedia, NNDB, Fashion Model Directory and MusicBrainz, as well as data contributed by its users.

Some knowledge bases that are used for a commercial service include Knowledge Vault [19] developed by Google and Probase [20] constructed by Microsoft. Knowledge Vault is a web-scale probabilistic knowledge base that combines extractions from web content with prior knowledge derived from existing knowledge repositories, which employs supervised machine learning methods for fusing these distinct information sources and features a probabilistic inference system that computes calibrated probabilities of fact correctness. Probase is designed to model the world as a model, and extracts knowledge from web fitting to their model.

DBpedia is one of popular structured knowledge bases [8][9]. It is constructed by extracting structured content from the information created as part of the Wikipedia data [9]. DBpedia allows users to semantically query relationships and properties associated with Wikipedia resources, including links to other related datasets.

In the areas of entertainment, there are well-known data sources: The Internet Movie Database (IMDB) is an online database of information related to films, television programs, and video games, including cast, production crew, fictional characters, biographies, plot summaries, trivia and reviews [10][11]. As of September 2015, IMDB had approximately 3.4 million titles (includes episodes) and 6.7 million personalities in its database, as well as 60 million registered

users*. On the other hand, Musicbrainz is an open content music database, and provides a structured open online database for music [12]. As of 25 October 2015, MusicBrainz contained information about roughly one million artists, 1.5 million releases, and 15 million recordings†. The Linkedbrainz provides a structured and semantic version of Musicbrainz data by mapping with between the Musicbrainz schema and an ontology model.

Although some of existing knowledge bases are already connected at a semantic level using linked data technologies, domain-specific data sources are also constructed in a graph knowledge format to interlink existing knowledge. For achieving this, ontological data model is essential, and knowledge-processing techniques play an important role to allow data to be collected and transformed into a graph knowledge format. Knowledge extraction from text employs the distant supervision method [13] to extract relations between recognized entities, which handle a task in two stages. One is the pattern training based on seed data and training corpus; the other is the extraction work applying the patterns on sentences. Natural language processing technologies, such as named entity recognition [14], dependency parsing [15][16][17], and co-reference resolution [18], are mandatory for this work. Finally, the extracted facts are transformed into triple format referencing to the designed ontology. Compared with this work, our solution will focus on fusing the multilingual knowledge, and provide unified entry for knowledge accessing.

3. A knowledge base for entertainment

In this section, we describe a process for constructing a knowledge base, which is comprised of ontology modelling, data collection and knowledge transformation of large-scale data.

3.1 Knowledge model of entertainment domain

A proposed ontology model is to provide comprehensive expressivity for various domains such as person or celebrities, TV programmes, time, movies, locations and music. Each domain is described in an individual class and those classes are semantic relations with others. For example, the Music class is linked to the class Person for expressing a set of relations such as singer, composer or producer. Figure 1 illustrates core classes of entertainment ontology, and shows some relationships among other classes. The Person class plays an important role to link other classes, and the Time and Location classes can be used for representing contextual information by connecting other classes.

Note that this knowledge model reuses a set of existing vocabularies to leverage connectivity of distributed data sources. For example, we do not create new classes for person, programme, movie and music. The following describes vocabularies that we reuse in our ontology model:

- Friend of a friend ontology [24]: machine-readable ontology describing personal profiles, their activities and their relationships.
- Music ontology [25]: concepts and properties for describing music domains (i.e. artists, albums, tracks, performance, etc.).
- Programme ontology [26]: a simple vocabulary for television programmes that covers brands, series, seasons, episodes and broadcast events and services.
- Dublin core metadata [27]: a set of metadata elements for cataloguing library items and other electronic resources.

Table 1 summarises core classes and their properties. As explained, some of properties are reused from existing vocabularies. The vocabularies that are defined by our own purpose are also described, and in this case, a namespace is defined as "sage".

3.2 Knowledge construction framework

To construct an entertainment knowledge base, we develop a framework, which is to handle structured and unstructured data from large-scale and heterogeneous data sources, as shown in Figure 2. The data collection component has several modules for aggregating a set of data sources such as a crawler for websites and a harvester for data dumps. A set of data is collected and stored in raw data storages without any further processing.

Then, structured data is generated by using the knowledge extraction module, which is to extract knowledge - a set of entities, comprehensive values of each entity and their relations with others - from multiple heterogeneous sources, including HTML pages, web tables and plain texts. We conduct some different ways of knowledge extraction as follows.

- HTML should consider a processing of unstructured and plain text and semantic mark-up, including RDFa, Microdata and Microformats, which combines several ontologies.
- Web table is also a valuable part for extraction. We trained a classification model based on decision tree to detect the tables. And considering that not all tables are filled with useful information, we develop a filter to remove the useless tables, which do not contain relational knowledge.
- Wikipedia data is also extracted to enrich the knowledge base. This task recognises info boxes in the pages, and parses the boxes to extract facts in triple format. Furthermore, we also consider multilingual support and real-time updating strategy for this task.

* <http://www.imdb.com/stats>

† <https://musicbrainz.org/statistics>

Table 1 Core classes and its main properties

Entity Class	Property	Range
Person (foaf:Person)	foaf:name	rdfs:Literal
	sage:person_imdb_image	-
	sage:person_imdb_link	-
	sage:birth_year	-
	sage:birth_monthday	-
	sage:death_year	-
	sage:death_monthday	-
	sage:height	xsd:string
Programme (po:Brand)	dc:title	-
	sage:imdb_rating	xsd:float
	sage:summary	xsd:string
	sage:genre	-
	dc:language	-
	sage:release_date	xsd:date
	sage:runtime	xsd:duration
	sage:shooting_location	xsd:string
	po:series	po:Series
	sage:director	foaf:Agent
	sage:writer	foaf:Agent
	sage:producer	foaf:Person
Movie (sage:Movie)	dc:title	-
	sage:imdb_image	-
	sage:imdb_rating	xsd:float
	sage:summary	xsd:string
	sage:genre	-
	dc:language	-
	sage:release_date	xsd:date
	sage:runtime	xsd:duration
	sage:shooting_location	xsd:string
	sage:director	foaf:Agent
	sage:writer	foaf:Agent
	sage:producer	foaf:Person
sage:sound_track	mo:MusicalManifestation	
Music (mo:Release)	mo:musicbrainz	foaf:Document
	foaf:maker	foaf:Agent
	dc:title	-
	ov:country	xsd:string
	dc:language	-
	dc:description	-
	mo:release_type	mo:ReleaseType
	mo:producer	foaf:Agent
	mo:publisher	foaf:Agent
	sage:composer	foaf:Agent
	sage:lyricist	sage:Entertainer
	mo:label	mo:Label
	mo:genre	mo:Genre
	sage:of_theme	sage:Theme
sage:of_mood	sage:Mood	
Time	sage:screen_date	xsd:date
	sage:screen_time	-

(sage:Showtimes)	xsd:date	-
Location (sage:Place)	geo:lat	-
	geo:lg	-
	sage:of_city	xsd:string
	sage:of_country	xsd:string
	sage:of_region	xsd:string

The extracted data should be refined to identify appropriate values of entities and attributes. Using string similarity methods, various typos and different formats are refined and replaced with consistent values. Then, candidates for entities and their descriptions are extracted.

A large volume of data is extracted from various data sources using the knowledge extraction framework, and then it should be transformed into a graph format using the knowledge model. For example, all of extracted values are transformed into a triple (i.e. subject, predicate, and object) in RDF. In particular, values extracted from tables are transformed by using data cube ontology for statistical web tables in order to obtain numerical knowledge.

After refining and extracting proper values for both entities and properties, we construct our knowledge base using the proposed knowledge model. It contains approximately 35 million entities and 850 million facts, ranging from six main subjects such as music, movie and TV, POI, dining, sports and locations as shown in Table 2.

3.3 Entity resolution and knowledge consolidation

Although the constructed knowledge is used for a fundamental source, it is also necessary to interlink existing data sources both structural and semantic data. In fact, various data sources can be used for commercial end-user services.

Figure 2 Knowledge construction framework

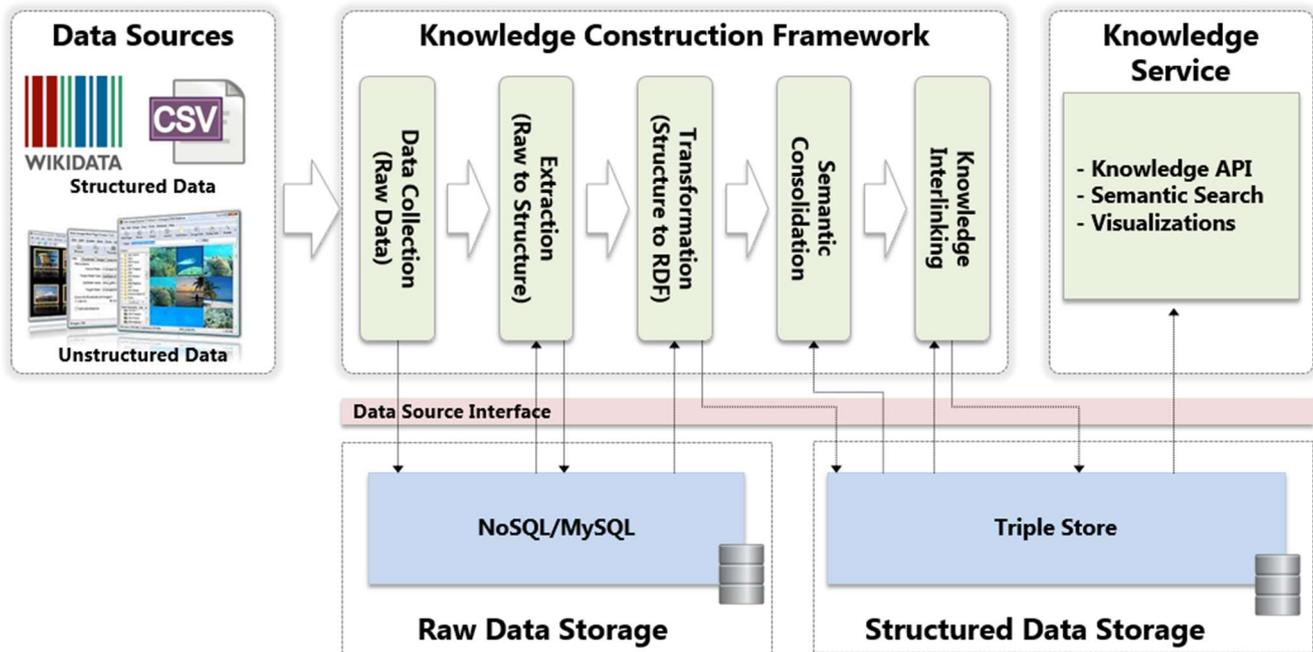


Table 2 Statistics of extracted data

Domain	Entity	A number of entities	Domain	Entity	A number of entities
Music	Artist	801,785	Dining	Restaurant	1,213,604
	Album	1,218,191		Menu	6,917,281
	Song	15,534,573	Sports	Team	5,988
	Cover art	318,614		Player	98,972
	Musical work	437,671		Relation	109,470
Movie/TV	Movie	362,657	Location	Restaurant	1,199,477
	TV brand	88,728		Nightclub	82,210
	TV episode	1,674,003		Shopping	444,958
	Person	2,046,214		Health	323,585
POI	Location	2,236,696		Hotel	9,251
Total number of entities		35,123,928			

Entity resolution is a task of disambiguating manifestations of real world entities [28]. There would be different ways of addressing same people, movies or particular objects. For example, a title of movies is often used with different names (e.g. "Men in black" vs. "MIB"). To integrate instances with their properties (values), this task is essential to define a unique identifier for resolving a set of entities among multiple data sources.

Figure 3 illustrates a process of entity resolution among multiple data sources. Methods that perform this task are categorised as blocking, iterative, or learning. Block-based approaches group together, in the same block, entity descriptions that are close to each other, denoting possible matches. To accomplish that, we define mapping rules (i.e. blocking keys) that is based on which the descriptions are placed into blocks. To identify same movies, we use a set of

mapping properties, including *dc:title*, *sage:release_date* and *sage:director*, while we use *po:series* and *po:position* properties for disambiguating TV programmes.

For computing similarity scores between two entities, we use the Jaro-Winkler distance, which the higher a distance score for two entities is, the more similar the string are [29]. The jaro similarity metric d_j for two given strings s_1 and s_2 is

$$d_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) & \text{otherwise} \end{cases}$$

where m is the number of matching characters and t is half the number of transpositions. Thus, two strings are

considered to be a match if and only if they are the same and they are t most at a distance from each other.

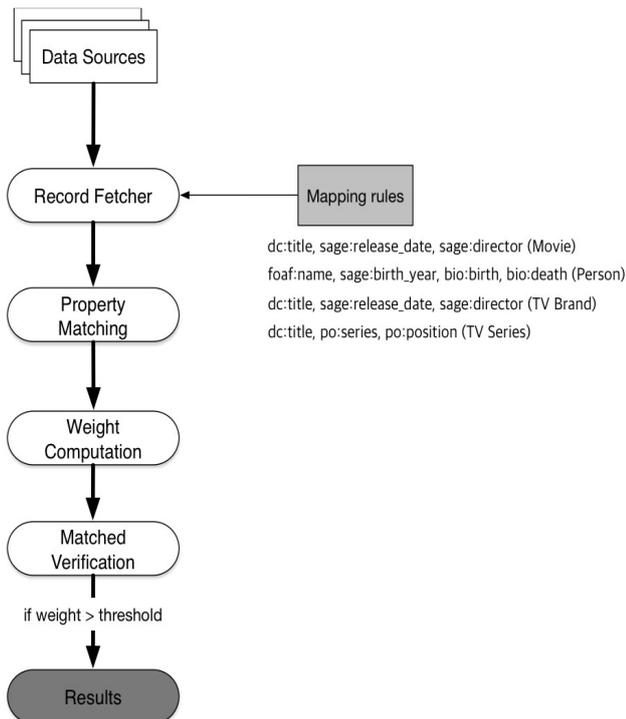
The Jaro-Winkler measure is an extension of the Jaro distance. Given two strings s_1 and s_2 , their jaro-Winkler distance d_w is:

$$d_w = d_j + (\ell p(1 - d_j))$$

where d_j is the Jaro distance for strings s_1 and s_2 , ℓ is the length of common prefix at the start of the string up to a maximum of 4 characters, and p is a constant scaling factor for how much the score is adjusted upwards for having common prefixes. p should not exceed 0.25, otherwise the distance can become larger than 1. The standard value for this constant in Winkler's work is $p = 0.1$.

We compute similarity measures of entities between the knowledge base that we construct and IMDB. In particular, we collect entities from IMDB: 422,916 entities for movie, 415,975 entities for TV series, 2,571,661 entities for TV episodes and 861,889 entities for person. For evaluating our algorithm, we use a random sampling, which extracts 1,000 linkage results. As a result, the extracted same entities obtain 98.5% accuracy. If results contain ambiguities after this step, we expand candidates such as release date, director, brand

Figure 3. An entity resolution process



name, name of series or date of birth and death. Note that we calculate values of *dc:title* with a high weight score (i.e. 0.9) in order to minimize an amount of candidates for mapping rules. After filtering out candidate sets, we apply a priority-based sequential similarity computation iteratively until our predefined threshold meets the condition. A mapping weight is defined by experts and experiments according to different

condition. Finally, the extracted same entities are represented in RDF using *owl:sameAs* property.

4. Using entertainment knowledge

In this section, we introduce a mobile application using the entertainment knowledge base, which allows users to discover relevant content across heterogeneous domains using the entertainment knowledge base.

In general, content recommendation would be highly depended on metadata of individual item [1][3] and user profiles such as preferences, demographical information or history of usage [4]. However, a means for an effective recommendation is often limited, when a set of available metadata of a single data source is not sufficient [2]. As introduced in the previous section, we constructed our own knowledge base for entertainment domains, and further this knowledge is interlinked other data sources. It is useful for searching relevant item for a certain keyword; for example, a search would be started from a primary data sources, and a search can be expanded when more relevant information is necessary throughout the interlinked data. Furthermore, the primary data does not provide associations of other content; we can investigate to search relevant items using interlinked data sources. Using these interlinked data sources, we can deliver related and recommended content to users at the right time. In fact, this feature is nothing new but the keys to differentiation are coverage of multiple domains.

In particular, this application provides a natural language interface to allow users to query in their everyday or natural language. Some of question-and-answer patterns are supported. For example, user can discover a following query: "How long is the movie Avengers?" Then the system returns the answer like "3 hours 15 minutes". Furthermore, this application provides a set of related content across different domains (e.g. movie, TV program and music). Figure 4 illustrates system architecture of our application. It is comprised of a client for end-user application and server-side components for question answer and knowledge bases. The following is detail about main components:

- User interface: this application allows users to discover and display a set of entertainment content on mobile environments.
- Question analysis: a question is transformed into a dependency structure by a robust parser, and question-pattern rules are applied to extract a type (what and how type) and content of the question. If no rule is applied, no type is extracted.
- Query construction: It discovers a query template based on the results of the question analysis, and creates a SPARQL query.
- Answer Generation: It generates an answer sentence using answering templates and knowledge entities from the entertainment knowledge base.
- Result Generation: It merges a set of answers for the question, a set of recommendation. A result can be displayed by computing a simple ranking algorithm.

highly flexible structure to interlink and expand other

Figure 4 A question answer application using the entertainment knowledge base

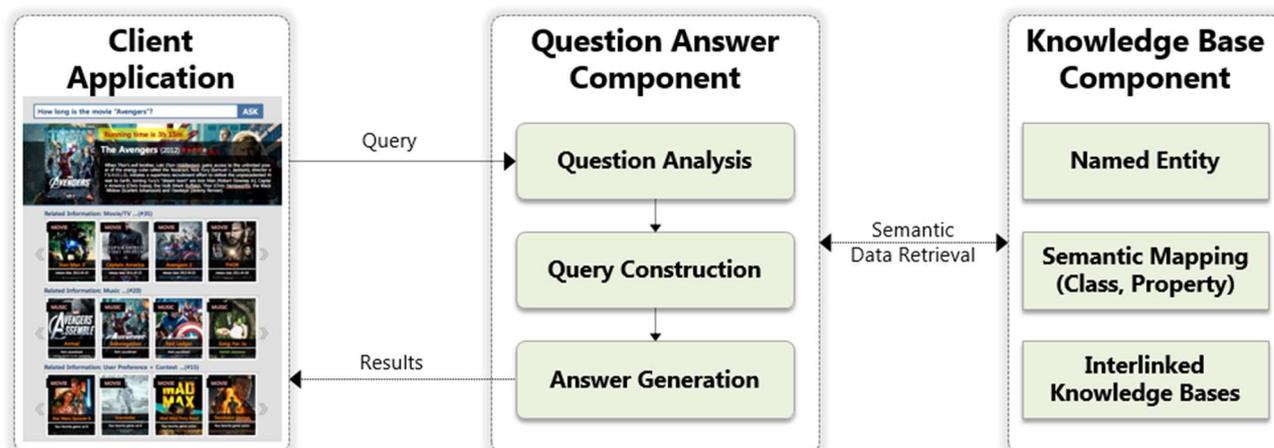


Figure 5 illustrates a mobile application, which contains examples of a movie and a musician, respectively. As explained, a dialog interface provides to discover some items, and some of question candidates are displayed by an auto-complete feature, when a pattern is matched. If a question is a type of answers, we provide the answer with details about the object: for example, user wants to know a running time of the movie “Avengers” in Figure 5. We give the answer (i.e. 3h 15m), and also provide description and rating about the item. In addition, a list of related items are provided simultaneously: the first panel provides a list of relevant movies and TV programs, which have same genres, directors, actors, and characters. We can find the other series of the selected item, and other sci-fi movies and TV programs. The second panel lists a set of associated songs with the Avengers. The last panel shows the results of personalized recommendation. A user can configure individual preferences on the application, including demographical profiles (birth date and place, and address), favorite genres and musicians, favorite actors, genres and directors. This information is used to analyze users’ personalization and recommendation, when an individual query is created.

5. Conclusion

This paper described a knowledge base for entertainment domains, which is created by integrating heterogeneous data sources. This knowledge contained more than 18.2 million entities and 328 million relations as facts. To construct this base, knowledge processing techniques were largely applied to extract, integrate, transform and interlink a number of entities and relations from heterogeneous data sources. Furthermore, a knowledge model for entertainment domains is developed to represent ontological semantics of individual content and its relations among others. This model has

ontology vocabularies such as book, sports or concert.

We also demonstrated a mobile application for content recommendation using the knowledge base. It allows users to a natural language interface to investigate an item that they want, by using a question & answering engine. A precise and relevant content item would be recommended to users by computing comprehensive metadata and its relations. In future, we will expand the proposed ontology model to other domains, and will develop mapping rules and ontology model for aligning other de facto model such as schema.org.

References

- [1] Choi, SM. et al. (2012) A movie recommendation algorithm based on genre correlations. *Expert Systems with Applications* 39(9):8079-8085
- [2] Ostuni, VC. et al.(2013) Mobile Movie Recommendations with Linked Data. *Availability, Reliability, and Security in Information Systems and HCI. LNCS 8127:400-415*
- [3] Mirizzi, R. et al. (2012) Movie Recommendation with DBpedia. *Proceedings of the Italian Information Retrieval Workshop (IIR), Bari, January 26*
- [4] Kaminskas, M. et al. (2014) Knowledge-based identification of music suited for places of interest. *Information Technology & Tourism*14(1):73-95
- [5] Bollacker, K. et al. (2008) Freebase: a collaboratively created graph database for structuring human knowledge. *Proceedings of the 2008 ACM SIGMOD, Vancouver, June 9-12, 1247-1250.*
- [6] Dong, L. et al. (2015) Question Answering over Freebase with Multi-Column Convolutional Neural Networks. *The 53rd Annual Meeting of the Association for Computer Linguistics, 260-269 .*
- [7] Raimond, Y. et al. (2012) Case Study: Use of Semantic Web Technologies on the BBC Web Sites, <http://www.w3.org/2001/sw/sweo/public/UseCases/BBC/>
- [8] Bizer, C. et al. (2009) Linked Data –The Story So Far. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5(3):205-227.

- [9] Bizer, C. et al. (2009) DBpedia – A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web* 7(3):154-165.
- [10] Avancha, S. et al. (2001) Design of Ontology for The Internet Movie Database (IMDB). Project Report CMSC 771 Spring 2001.
- [11] Semeraro, G. et al. (2009) Knowledge infusion into content-based recommender systems. *Proceedings of the third ACM conference on Recommender systems*, New York, October 22-25, 301-304
- [12] Swartz, A. (2002) Musicbrainz: A semantic web service. *Journals of Intelligent Systems*, 17:76-77.
- [13] Roth, B. et al. (2013) Effective Slot Filling Based on Shallow Distant Supervision Methods. *Proceedings of the Sixth Text Analysis Conference (TAC)*, Maryland, November 18-19.
- [14] Lafferty, A. et al. (2001) Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning (ICML)*, Williamstown, 282–289.
- [15] Bikel, DM. (2004) On the Parameter Space of Generative Lexicalized Statistical Parsing Models. Doctoral dissertation. University of Pennsylvania, Philadelphia.
- [16] Klein, D. et al. (2003) Accurate Unlexicalized Parsing. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics* 1:423-430.
- [17] Zhu, M. et al. (2013) Fast and Accurate Shift-Reduce Constituent Parsing. *Proceedings of the Annual Meeting of the Association of Computational Linguistics*.13-10
- [18] Lee, H. et al. (2013) Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Association for Computational Linguistics* 39(4):885-916.
- [19] Dong, X. et al. (2014) Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. *Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, New York, 601-610.
- [20] Wang, Z. et al. (2010) Probase: a Universal Knowledge Base for Semantic Search. Microsoft Research Asia.
- [21] Ferrara, A. et al. (2011) Data linking for the semantic web. *International Journal on Semantic Web and Information Systems* 7(3):46-76.
- [22] Jain, P. et al. (2010) Ontology Alignment for Linked Open Data. *Proceedings of the international Semantic Web Conference*. LNCS 6496:402-417.
- [23] Waitelonis, J. et al. (2012) Towards exploratory video search using linked data. *Journals of the Multimedia Tools and Applications* 59:645-672.
- [24] Graves, M. et al. (2007) FOAF: Connecting People on the Semantic Web. *Cataloging & Classification Quarterly* 43(3-4):191-202.
- [25] Raimond, Y. (2007) The Music Ontology, *Proceedings of the International Conference on Music Information Retrieval*: 417-422
- [26] Kobilarov, G. et al. (2009) Media Meets Semantic Web - How the BBC Uses DBpedia and Linked Data to Make Connections. *The Semantic Web: Research and Applications*. LNCS 5554:723-737.
- [27] Beckett, D. et al. (2002) Expressing Simple Dublin Core in RDF/XML. Technical report, Dublin Core Metadata Initiative.
- [28] Kopcke, H. et al. (2010) Evaluation of entity resolution approaches on real-world match problems. *Proceedings of the Very Large DataBase(VLDB) Endowment* 3(1-2):484-493
- [29] Islam, A. et al. (2008) Semantic Text Similarity Using Corpus-Based Word Similarity and String Similarity. *ACM Transactions on Knowledge Discovery from Data(TKDD)* 2(2): Article 10.

How long is the movie "Avengers"? **ASK**

Running time is 3h 15m

The Avengers (2012) ★★★★★ (8.1)

When Thor's evil brother, Loki (Tom Hiddleston), gains access to the unlimited power of the energy cube called the Tesseract, Nick Fury (Samuel L. Jackson), director of S.H.I.E.L.D., initiates a superhero recruitment effort to defeat the unprecedented threat to Earth. Joining Fury's "dream team" are Iron Man (Robert Downey Jr.), Captain America (Chris Evans), the Hulk (Mark Ruffalo), Thor (Chris Hemsworth), the Black Widow (Scarlett Johansson) and Hawkeye (Jeremy Renner).

Related Information: Movie/TV ...(#35)

- MOVIE: Iron Man 3 (release date: 2012-04-25)
- MOVIE: Captain America (release date: 2011-07-22)
- MOVIE: Avengers 2 (release date: 2015-04-23)
- MOVIE: THOR (release date: 2011-04-28)

Related Information: Music ...(#20)

- MUSIC: AVENGERS ASSEMBLE - Arrival (from soundtrack)
- MUSIC: AVENGERS ASSEMBLE - Interrogation (from soundtrack)
- MUSIC: AVENGERS ASSEMBLE - Red Ledger (from soundtrack)
- MUSIC: Song For Jo (Scarlett Johansson)

Related Information: User Preference + Context ...(#15)

- MOVIE: Star Wars: Episode II (Your favorite genre: sci-fi)
- MOVIE: Interstellar (Your favorite genre: sci-fi)
- MOVIE: Mad Max: Fury Road (Your favorite genre: action)
- MOVIE: Terminator Genisys (Your favorite genre: action)

Who sings "She's Not Me"? **ASK**

Madonna

- Gender: Female
- Country: United States
- Type: Person, Singer, Entertainer
- Genre: Electronic, Pop, Dance, Electropop

Related Information: Movie/TV ...(#15)

- TV: MTV Video Music Awards 2009 (Madonna were starring)
- TV: Hope for Haiti Now (Madonna were starring)
- MOVIE: Evita (Madonna were starring)
- MOVIE: W.E. (Madonna were directing)

Related Information: Music ...(#30)

- MUSIC: Hard Candy (Madonna sing)
- MUSIC: Sticky & Sweet Tour (Madonna sing)
- MUSIC: I'm Going to Tell You a Secret (Madonna sing)
- MUSIC: Like a Virgin (Madonna sing)

Related Information: Event ...(#8)

- EVENT: Cologne, Germany (Madonna's Concert (11/04))
- EVENT: O2 arena, Czech (Madonna's Concert (11/07))
- EVENT: Tele2 Arena (Madonna's Concert (11/14))
- EVENT: Palau Sant Jordi (Madonna's Concert (11/24))

Figure 5 A question answer based mobile application. This application allows users to find out some items by natural language in English. The left shows an example to have a question and answer in some subjects of film (i.e. a running time of Avengers), while the Note that this application provides relevant items such as film, music and event using the entertainment knowledge base.