

Interactive Exploration of the Movie DB on a Semantical Level

Thorsten Liebig*
Ulm University, Germany

Olaf Noppens†
Ulm University, Germany

Timo Weithöner‡
Ulm University, Germany

ABSTRACT

The IMDb can be seen as an ontology made of a schema and a huge network of individuals. This allows to reason about movie data and to define filters in terms of declarative descriptions. We believe that many interesting queries about movies can be answered by interactive visual exploration utilizing browsing primitives such as aggregated club views, selective expansions, or drag-n-drop filters.

1 MOTIVATION AND APPROACH

The Internet Movie Database (IMDb) is a large collection of movie information mainly compiled from user submissions. The core of an IMDb movie record consists of data about involved persons (e. g. director, cast), about the movie itself (genre, locations, awards) as well as additional details (rating, length, etc.). The two former collections have some similarities to a social network. In this regard the movie network is a structure made of different types of nodes (actors, movies, genres, etc.) that are tied by one or more specific kinds of relations (movie appearances, oscar achievements etc.).

In Knowledge Representation (KR), such a vocabulary of named nodes (typically called concepts) and relationships (called properties) along with a formal semantics is called an ontology. Having a formal semantics allows to precisely characterize the concepts and properties of the focused domain (e. g. by defining a property as functional or a concept by a logical expression). This enables to make implicit information within the ontology explicitly available with the help of a reasoning engine.

Our approach combines logical reasoning over an enriched ontological model of the IMDb with an interactive, semantically guided investigation of the manifold interconnected movie network in order to visualize and understand the larger structure of the data or to answer specific tasks. We propose an user-driven exploration strategy, making use of animated expansion steps, clustering techniques, and different levels of detail or abstraction views. During exploration, any expansion set may be narrowed by drag-n-drop of restricting criteria. Instant reasoning feedback is employed for automatic instance classification or to check for conflicting filtering criteria. This is done with help of a specially enhanced version of our ontology browsing and authoring tool ONTOTRACK [1] linked to our high performance relational reasoner U2R2.

2 THE MOVIE ONTOLOGY

An ontology typically is divided into two parts: the schema part introduces concepts and properties and gives structure to them in terms of axioms using language constructs such as sub-concept, sub-property or logical operators. The assertional or data part defines individuals and concrete relationships between those individuals utilizing the concepts and properties of the schema.

*e-mail: thorsten.liebig@uni-ulm.de

†e-mail: olaf.noppens@uni-ulm.de

‡e-mail: timo.weithoener@uni-ulm.de

2.1 Movie Schema

Within the movie domain our schema obviously contains terms such as `Movie`, `Person`, `Female`, or `Genre` and properties like `hasGenre`, `directedBy`, `cast`, or `leadingActorOscar`. The latter relates a movie to a person and is defined as a sub-property of `cast` as well as functional, which means that one particular movie can only have at most one `leadingActorOscar` (other types of oscar awards are still allowed of course). In addition, most properties have an inverse counterpart. For instance, `appearsIn` is the inverse of `cast`.

Furthermore, we have defined some advanced concepts of obvious interest within this domain. For example, an `OscarMovie` is defined as a movie who is related to a person via at least one of the `oscarAward` properties (Fig. 2 shows this definition in a logical KR notation). Moreover, a `Performer` is defined as either an `Actor` or `Actress`. A `PerformerOfAnOscarMovie` in turn is a performer who appears in some `OscarMovie`.

2.2 Movie Data

The corresponding assertional part covers all the data about concrete movies, persons, received oscars etc. For instance, in 2005 Cate Blanchett received an award for the best supporting actress in the “The Aviator”. Therefore, the supportingActressOscar property holds between the individuals `Aviator.The` and `Cate.Blanchett`. As a result, `Aviator.The` will be derived as an instance of `OscarMovie` by our reasoning engine.

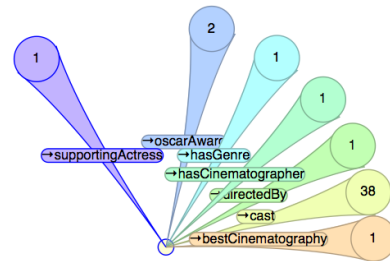


Figure 1: Clickable preview clubs for expanding fillers.

The Web Ontology Language (OWL) has been used as representation format for both parts of the movie ontology. The movie schema with its roughly 50 concepts and 20 properties has been modeled manually following the IMDb record structure. The data part was mapped into OWL syntax by a XSL transformation. To show extensibility wrt. the underlying date we added a `posterURI` to our model storing the corresponding poster URLs which we gathered from the IMDb web site.

3 INTERACTIVE VISUAL EXPLORATION OF THE IMDb

We utilize an adapted version of our ONTOTRACK ontology IDE for exploring the movie data in the following way. The schema authoring component can be used to browse through the concepts and properties as well as to specify additional concepts whenever needed. The data visualization component allows a user-directed exploration of interrelated individuals [2]. The user can start by dragging an individual from a list of all individuals onto the data

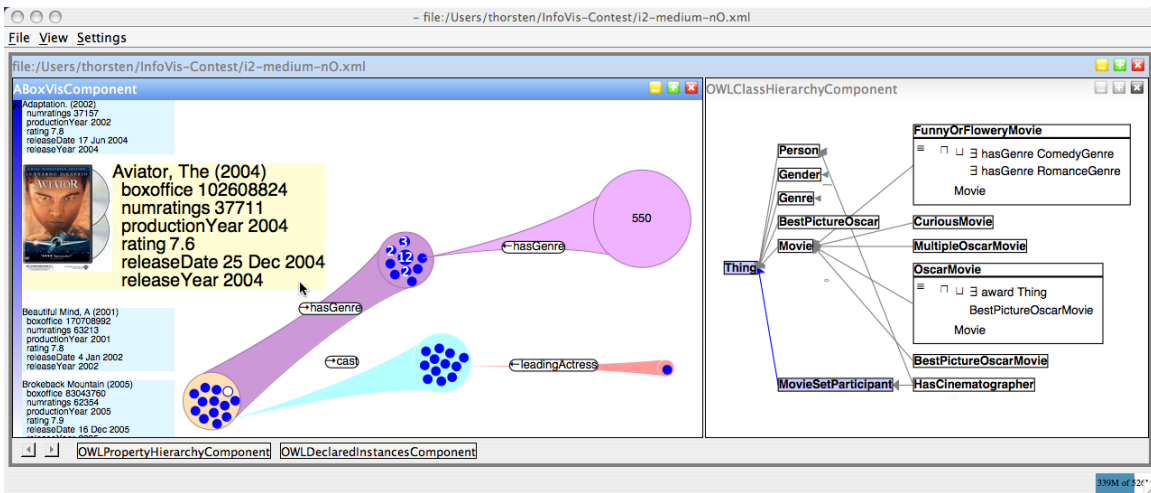


Figure 2: Application screen shot showing data pane on the left and schema pane on the right.

pane. This individual as well as all others are depicted as circles within the data pane. Now the user can interactively exploit the related individuals (property fillers) in a step-wise fashion. On mouse click a preview of clubs will show those properties which have fillers and their quantities with respect to the originating one. Figure 1 shows the preview for the movie “The Aviator”. One can see that there are 38 fillers of the cast property, 2 fillers of the award property etc. A preview club will expand in an animated fashion when selected by mouse. After expansion, all fillers of this property are grouped within a so-called *property filler cluster* which will be drawn as a colored club (cp. Fig. 3). Now, the user can select a new or the same individual as the source of further property expansions.

To distinguish the various properties different colors are used. Alternatively it is also possible to expand the fillers of a whole cluster with respect to a property as shown in Figure 3. This can be done by choosing the club itself as expansion source. The emerging filler cluster then contains the union of all fillers and is depicted by a broader interconnection. For instance, to reveal all genres of a

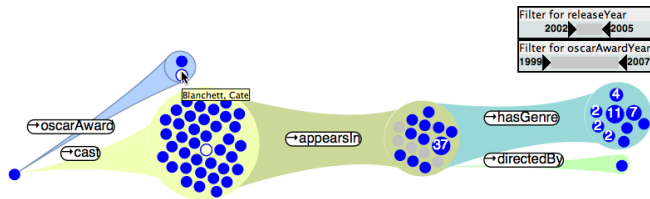


Figure 3: Expanded chain of clubs with release date constraints.

set of movies one simply can expand the union of the *hasGenre* property as shown within Figure 3 with one single mouse operation.

Typically each genre is referenced from a couple of movies. In order to visualize these quantities the diameters of the genre circles will scale proportionally with the number of related movies in the predecessor club. In case there is more than one source individual we also write the number of those on the circles. Whenever expanding a surjective property this technique allows to grasp the distribution of relationships to fillers. E. g. the accumulation of associated genres within a set of movies is depicted as shown in Figure 3.

A list with details for clustered individuals is optionally available. The list content will instantly switch to those individual set the user is hovering with the mouse. For instance, when listing movies their ratings, release and production date as well as movie

poster is shown on the left hand side of the view pane. This list can be scrolled and provides a mouse over magnifying effect.

A second dimension of exploration is given by drag-n-drop operations with concepts from the schema. Dropping a concept onto the background of the data pane results in a new initial set of root individuals (the instances of this concept). A concept can also be used as a local filter when dropped onto an existing cluster. For instance, any set of movies can be restricted to oscar movies simply by dragging the *OscarMovies* concept onto this cluster. Non oscar movies will then disappear from this cluster. This is a powerful selection mechanism since concepts can be defined on demand during exploration. The underlying reasoning engine supports the user by instant classification of individuals as well as in providing feedback about nonsensical restrictions (e. g. restricting a set of actresses to male persons inevitably would lead to an empty set).

A local filter, however, may affect adjunctive or prospective clusters. Consequently, a filter will propagate its restrictions. Therefore, when restricting a set of movies to oscar movies each successive club will adapt its filler sets only to those individuals actually related to oscar movies.

Since time is an important variable within the movie domain we also enhanced our tool with user adjustable time sliders. A time slider allows to reveal the correlation of individuals to particular time points or intervals. For example, each movie has fillers for the properties *releaseYear* and *productionYear*. The slider for the latter allows to specify any arbitrary interval of years between 1999 and 2007. All those movies which do not fall into the actual interval will be instantly colored grey. Similar could be done with other concrete domains such as ratings or box office actuals.

3.1 Data Specific Tasks

The proposed approach and our system is best suited for answering qualitative queries, such as the first example from the contest web page (“Do Best Actress Oscar winners tend to come from certain genres?”). Even more complicated queries are easily answered such as “How many actresses who sometime played in a movie which received an oscar also appeared in an adult movie?”.

REFERENCES

- [1] T. Liebig and O. Noppens. ONTOTRACK: A semantic approach for ontology authoring. *Journal of Web Semantics*, 3(2):116 – 131, 2005.
- [2] O. Noppens and T. Liebig. Understanding Large Volumes of Interconnected Individuals by Visual Exploration. In *Proc. of the 4th Europ. SW Conf. (ESWC 2007)*, pages 799–808, Innsbruck, Austria, 2007.