

## Introduction to Working with Semantic Web Analysis Graphs

Step	Action and Explanation
1)	<p>Systems for Online Analytical Processing (or OLAP) support business analysts in their decision-making processes by allowing them to view the data at different granularities. In order to be accessible to OLAP, the data must be organized according to a multidimensional model. A multidimensional model consists of facts, which are the subjects of the analysis and quantified by measures. A multidimensional model also consists of hierarchically-organized dimensions which allow for measure aggregation.</p> <p>RDF is a standard data model for interchanging data on the web, where resources are identified by Uniform Resource Identifiers (or URIs). RDF data are in the form of triples of subject, predicate, and object. External RDF data, such as DBpedia and Wikidata, can be an important source of knowledge for OLAP analysis.</p> <p>You are now getting an introduction to semantic web analysis graphs, which allow you to pose OLAP queries against RDF data sources without using the SPARQL query language which is typically used for querying RDF data. In the course of this introduction, you will receive instructions that you should follow step-by-step. After the introduction, you should complete a couple of tasks and fill out a questionnaire.</p>
2)	<p>On the sheets handed out to you, you will find the address of a web application. Please open the site using your Google Chrome (!) browser. On the webpage, please choose “Films” and click “Select”.</p> <p>The nodes of an analysis graph are analysis situations, each representing a multidimensional query. The edges of an analysis graph are navigation steps, each representing one or more OLAP operations that transform the source analysis situation into the target analysis situation, thereby plotting a recommended course for the analysis. Analysis graph schemas may contain variables to be bound to concrete values during instantiation, which we will show later.</p>
3)	<p>In the top part of the webpage, note that the “Analysis” tab is selected. Note also that the “Analysis” tab has two panes (or sides). The pane on the right-hand side has the tabs “Graph”, “Results”, and “SPARQL”. The “Graph” tab is selected by default in the beginning and contains the analysis graph view where a recommended course for the analysis is displayed. The left-hand side is the details panel and shows the details of the elements that are selected in the analysis graph view.</p> <p>Take a look at the analysis graph in the graph panel on the right-hand side of the webpage. Analysis situations are the circles. Navigation steps are the directed arcs – the arrows – connecting them. Note the names of the analysis situations which are written next to the circles. Note also the names of the navigation steps which are the labels of the arrows.</p> <p>In the graph view on the right-hand side, zoom in and out using the mouse wheel. Click in an empty area of the graph view and hold the mouse button to drag the graph. During the study, use this feature to focus on a specific part of the graph.</p> <p>We now want to use the <code>AverageDurationOfFilms</code> analysis situation. Hover over the <code>AverageDurationOfFilms</code> analysis situation and see that its size is enlarged, the whole analysis graph fades a little, and a summary of the analysis situation is displayed in a black rectangle. Click on the <code>AverageDurationOfFilms</code> analysis situation and hold the button to drag it to a position of your choice. Note that clicking the <code>AverageDurationOfFilms</code> analysis situation causes it to be selected. Note that after the selection of the <code>AverageDurationOfFilms</code> analysis situation, its circle becomes checkered, its size is enlarged, and incoming and outgoing navigation steps are highlighted in blue. Notice the details pane of the analysis situation at the left-hand side of the webpage. Note that the details are organized in four sections: “Summary”, “Further Steps”, “Measures”, and “Dimensions”.</p> <p>Notice that the <code>restrictGenderOfDirector</code> navigation step is an outgoing navigation from the <code>AverageDurationOfFilms</code> analysis situation. Hover over the <code>restrictGenderOfDirector</code> navigation step and notice that the whole analysis graph fades a little and that a summary of the navigation step is displayed in a black rectangle. Click on the outgoing <code>restrictGenderOfDirector</code> navigation step, then see how the navigation step arrow is textured and enlarged and its source and target analysis situations are highlighted in blue.</p>

	<p>See the details pane of the navigation step at the left-hand side of the webpage. The organization of the details of the navigation step is as follows: The details are split between “Summary” and “Operations”. Note the source and target analysis situations <code>AverageDurationOfFilms</code> and <code>AverageDurationOfFilmsWithDirectorOfSelectedGender</code> highlighted in blue in the graph view on the right-hand side pane. Note also that the arrow representing the navigation step indicates the direction, i.e., from the source analysis situation <code>AverageDurationOfFilms</code> to the target analysis situation <code>AverageDurationOfFilmsWithDirectorOfSelectedGender</code>. Now right-click on the <code>restrictGenderOfDirector</code> navigation step and see the “Navigate to target analysis situation” option in the displayed menu.</p> <p>Note that you can search for analysis situations and navigation steps by name. In the search area above the analysis graph panel, write “<code>AverageDurationOfFilms</code>” and then click the “SEARCH” button. Note that the name of the <code>AverageDurationOfFilms</code> analysis situation is highlighted while the rest of the graph fades for an instant.</p>
<p>4)</p>	<p>Click the “Schema” tab and take a look at the employed multidimensional schema. This schema allows to perform analyses of the duration of films. The original data are in RDF format and are freely accessible via a SPARQL endpoint over the internet. The queries formulated using the analysis situations are rewritten internally into SPARQL queries.</p> <p>Film is the fact class. <code>Origin</code>, <code>Direction</code>, <code>ReleaseDate</code>, and <code>Content</code> are the dimensions. The <code>Origin</code> dimension describes the country of origin of the film and has levels of granularity <code>Country</code> and <code>Continent</code>, where <code>Country</code> rolls up to <code>Continent</code>. In the <code>Direction</code> dimension, <code>Director</code> rolls up to <code>Gender</code>. The <code>ReleaseDate</code> dimension models the release date of the film and has three levels of granularity, namely <code>Date</code>, <code>Month</code>, and <code>Year</code>. The <code>Content</code> dimension has no hierarchies. In each dimension there is an implicit <code>All</code> level that has a single value.</p> <p><code>Duration</code> is the measure and it refers to the runtime of the film (in minutes). The aggregation function “AVG” is used with this measure. For example, the average duration of British films in October 2010. The measures can be rolled up along the dimension hierarchies, e.g., <code>Duration</code> is available by <code>Country</code> and <code>Month</code>, and can then be calculated by <code>Continent</code> and <code>Year</code>.</p> <p>Click the “Help” tab. Allow pop-up windows if asked to and see the detailed description of the user interface items and how to use them. Please refer to the help if anything is unclear during the study.</p> <p>Now return to the main window and click the “Graph” tab in the top part of the right-hand side pane to go back to the analysis graph view.</p>
<p>5)</p>	<p>Next, notice the “New” tab in the top part of the page on the left-hand side. Now click this tab to go back to the first webpage to select the analysis graph and start over. The page will prompt for confirmation. Click “OK” to be taken back to the start. Then select the <code>Films</code> analysis graph again and click on “SELECT”. You can click this tab to restart the program at any time, which will cause your progress to be lost.</p> <p>Click on the <code>AverageDurationOfFilms</code> analysis situation to select it. This analysis situation calculates the average <code>Duration</code> of films. Additional parameters may be specified by the user by binding values to certain variables. In the details pane, note the “Summary” section, which shows a brief overview of the selected analysis situation. Note also the “Further Steps” section, which recommends further steps for the analysis. Click the section label “Further Steps” to expand the section. Note the two available steps. Click again on the “Further Steps” section label to collapse the section. Other sections can be expanded and collapsed in the same way. Hover over the information icon indicated by the letter “i” beside the “Further Steps” label. Note the displayed information on the “Further Steps” section. You can hover over the information icon to get more information about the elements. The measures section labeled “Measures” lists the displayed measures in this situation. This particular analysis situation has a single measure. Note that the AVG aggregation operator is applied to the <code>Duration</code> measure. Below the measures section, note the dimensions section of the analysis situation, which is labeled “Dimensions”. This analysis situation has a parameter for the <code>Origin</code> dimension which allows to restrict the <code>Continent</code> level to a specific value. This value must be a dimension member.</p>

	<p>Now we want to restrict the Continent to a specific value. See the field with the placeholder “dimension member” under the Continent label. Go to this empty field and start typing “Asia” - the autocomplete functionality will show you a list of suggestions; select “Asia”.</p> <p>The blue button labeled “RESULTS” at the very top of the details pane then allows you to perform the query and calculate the results. Click the “RESULTS” button to calculate the results. See the results in the “Results” tab on the left-hand side pane. Click twice on the AVG column header to sort the results in descending order according to this column.</p> <p>Now we want to edit the value we entered to restrict Continent and see how it affects the results. Click the remove button (labeled with an “x”) next to the field below Continent where you entered the value in order to clear the value. Start typing “Europe” and use the autocomplete functionality to cast your selection. To view the updated results, you can either click the blue “RESULTS” button, or click on the “Results” tab. Now click on the “Results” tab which causes the refreshing of the results using the updated parameters.</p> <p>Now remove again the restriction on Continent and click the “RESULTS” button again. Even though the parameter is not bound, you may still execute the query. In case a parameter is not bound, no restriction will be performed.</p> <p>Click on the “SPARQL” tab. This tab shows the SPARQL query that is automatically generated by the system for the analysis situation at hand. You can also edit the query. Now click the button labeled “EXECUTE” and allow pop-up windows if asked to. This will execute the query in a separate window. Note that normally you will NOT need this feature in the user study.</p> <p>Return to the main window and click the “Graph” tab in the top part of the right-hand side pane to go back to the analysis graph view.</p>
<p>6)</p>	<p>Click the <code>showResultsPerGenreAndRestrictYearOfRelease</code> navigation step which connects the <code>AverageDurationOfFilms</code> and <code>AverageDurationOfFilmsPerGenre</code> analysis situations. A navigation step represents a recommended course of analysis. Your selection made at the analysis situation where you start will be carried over to the following analysis situation.</p> <p>In the details pane, on the left-hand side of the webpage, notice the “Summary” label. Notice that in the graph view on the right-hand side of the webpage that this navigation step has the <code>AverageDurationOfFilms</code> analysis situation as its source and the <code>AverageDurationOfFilmsPerGenre</code> analysis situation as the target.</p> <p>In the details pane, note the section labeled “Operations” and take a look at the two operations of this navigation step. The first operation is <code>drillDownToGenre</code>. Notice the icon beside the operation which symbolizes its functionality. This operation performs a drill down to change the granularity in the Content dimension to the Genre level. Therefore, the target <code>AverageDurationOfFilmsPerGenre</code> analysis situation will have the Genre level as a granularity level in the Content dimension.</p> <p>Notice the second operation of this navigation step labeled <code>changeSliceConditionOverYear</code>. This operation restricts the Year of the <code>ReleaseDate</code> of the film to certain values using a condition that you may specify. Note that the condition is composed of two parts. The first part is a comparison operator. This operator is used to compare the values of the Year to a value of choice to be entered in the second field with the “value” placeholder. We will now restrict the Year to values greater than 2009, in order to get the average duration of films released after 2009. If you specify a comparison operator and a value to compare with and then follow the selected navigation step, the target analysis situation will have this restriction of the values of Year of the <code>ReleaseDate</code>.</p> <p>Let us now illustrate the workings of a navigation step. But first, click on the <code>AverageDurationOfFilmsPerGenre</code> analysis situation – the target analysis situation of the <code>showResultsPerGenreAndRestrictYearOfRelease</code> navigation step. This analysis situation returns the average Duration of films at the Genre granularity. Additional restrictions apply to the continent of origin of the film and to the year of release – variable parameters that you may specify. In this regard, remember that you may arbitrarily click into any analysis situation, specify the parameters and execute the query. But, navigating to an analysis situation via navigation step from your previous analysis situation has the advantage that</p>

	<p>the previous parameter bindings are carried over to the new situation, and you can follow the recommended course of analysis. Remember that the <b>Continent of Origin</b> and <b>Year of ReleaseDate</b> parameters of the selected analysis situation are currently not bound.</p> <p>Now click on the <b>AverageDurationOfFilms</b> analysis situation again. Restrict <b>Continent</b> to <b>Europe</b>. Next, click the <b>showResultsPerGenreAndRestrictYearOfRelease</b> navigation step. Restrict the <b>Year of ReleaseDate</b> to values greater than (&gt;) 2009. After having done so, click on the "NAVIGATE" button. The result then shows the average duration of films within the different genres, considering only films from Europe and released after 2009. Notice that the restriction on <b>Continent</b> has been carried over from the source situation. That way, you can navigate through the analysis graph from one analysis situation to the next. Click the "Graph" tab to get back to the analysis graph view in the pane on the right-hand side of the user interface.</p> <p>One final time, click on the <b>AverageDurationOfFilms</b> analysis situation and then expand the "Further Steps" label. Here, notice quick links to the outgoing navigation steps. Click on the "showResultsPerGenreAndRestrictYearOfRelease" to open the details of the navigation step of the same name.</p>
<b>7)</b>	<p>From now on, please follow the tasks that you find on your instruction sheets. Enter your answers in the fields provided on the instruction sheets. After completing a task, evaluate how difficult or easy you found the task using the rating scale at the end of each task. When you are done with all tasks, please complete the attached questionnaire. Please read the instruction sheets carefully. The tasks are to be solved individually.</p>