



KIM PLATFORM


WEB User Interface GUIDE

*The step by step guide
to exploring the KIM
functionalities through
the Web UI*

1 Table of Contents

1	Table of Contents	2
2	Introduction	3
2.1	Search & Navigation paradigms	3
2.2	Search results	4
2.3	Technology	4
3	Keyword Search Paradigm	5
3.1	Searching in the body of the document.....	5
3.2	Searching in the features of the document	5
3.3	Searching in the body and the features of the document.....	6
4	Conceptual Search & Navigation Paradigms	7
4.1	Structure	7
4.2	Patterns	11
4.3	Ontology navigation	12
5	Combined Search Paradigms	15
5.1	Facets.....	15
5.1.1	Keyword search	15
5.1.2	Co-occurrence of entities.....	16
5.1.3	Co-occurrence of entities and keywords	17
5.1.4	Configuration options	18
5.2	Timelines.....	18
5.2.1	Creating a timeline query.....	19
5.2.2	Timeline results	22
6	Search Results	23
6.1	Document Query Result	23
6.2	Entity Query Result.....	24
6.3	Timelines Result.....	25
7	Results Detail.....	27
7.1	Document Detail.....	27
7.2	Entity Description	28


2 Introduction

 latestnews is a web application for search and analysis over multiple sources of international and business news such as International Herald Tribune, BBC, NY Times, etc. It provides various search paradigms on top of a semantic database.

 latestnews allows you to:

- Find all available documents about an event, person, organization, etc.
- Define search by keywords, entity names, conceptual relations
- Explore links between entities occurring together in a text
- Analyze emerging trends about entities of interest over a period of time

2.1 Search & Navigation paradigms

In its essence,  latestnews performs three basic types of search – keyword (full-text search over the texts), conceptual (pattern search over the graph of the knowledge base), and combined.

Depending on the problem you want to solve, there are multiple Search Paradigms available.

Keyword (uses keywords/phrases)

- **Keyword:** allows you to combine words and phrases using Logical (Boolean) Operators (such as *AND*, *OR*, and *""*) to limit, widen, or define your search

Conceptual (uses entities from the knowledge base)

- **Structure:** can be used for two kinds of search tasks:
 - To look for entities through their names, part of their names, or entity type (Entity look-up)
 - To create patterns by choosing entity types, names, and relations between them. If you want to make you search even more specific, you can use additional entity characteristics
- **Patterns:** can be used for modeling frequently executed structured queries (Predefined patterns for searching entities)
- **Ontology:** can be used for exploration of the [PROTON Ontology](#) Classes of entities and their properties

Combined (uses keywords and/or entities)

- **Facets:** can be used for finding entities that co-occur with other entities and terms in the same context. In other words, it helps you find unknown entities


by defining the environment in which they appear (relations with other entities, dates, or terms)


- **Timelines:** can be used for research into how entity popularity changes through time in a given set of documents; trends of entity popularity

2.2 Search results

The results of the different search paradigms can be [documents](#), [entities](#), or [timelines](#).

2.3 Technology

 latestnews is based on [KIM 3](#) - Knowledge and Information Management, a platform for automatic semantic annotation, indexing, and retrieval of unstructured and semi-structured content.

 Note that, as the documents in the document repository may differ from the time this guide was created, the given examples may not work exactly as provided.

3 Keyword Search Paradigm

Boolean is a full-text search paradigm, providing a tool for searching in the document content, as well as in the document features and structure. It allows you to combine words using Logical Operators, such as *AND* & *OR*, to limit or widen your search. By looking for a given word that appears in the document content, you can retrieve a particular document or documents. This works like a classical search engine locating strings of characters in your set of documents.

On top of that, you can search for some of the additional information, available for each document in your document set. This additional information is called metadata and consists of different document features such as the source of the document, the date it was created, etc.

3.1 Searching in the body of the document

Body content text box

The **Body content** text box is the place where you enter the keyword(s) that define your search. The **Keyword** paradigm searches for these words in the body of your set of documents.

TASK: Find all documents that talk about a particular term, for example “petrol OR electric”.

TO DO: Type “petrol OR electric” in the **Body content** text box.

Body content:	<input type="text" value="petrol OR electric"/>
---------------	---


3.2 Searching in the features of the document

Feature and Feature value boxes

The **Feature** box is the place where you can select search criteria, based on the document features, i.e. the additional information available for the documents. These features can vary depending on the domain and the set of documents. In this case, as the corpus of documents consists of international news, some of the features are: the subject and author(s) of the news article; some key entities/key phrases that are selected as characteristic of the article; the title/subtitle; the original URL, where the article can be found; etc. The **Feature value** text box is where you type the value of you document feature.

TASK: Find all documents where a particular term appears in their title, for example the term “cars”.

TO DO: Select TITLE as your document feature from the **Feature** box and type “cars” in the **Feature value** box.



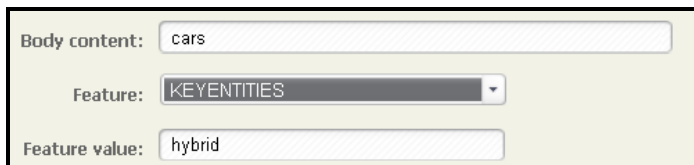
Body content:

3.3 Searching in the body and the features of the document

You can narrow your query even further by searching in both the document features and content.

TASK: Find all documents where a term and a document feature appear together, for example documents about the “cars” that have the word “*hybrid*” as KEYENTITY.

TO DO: Type “cars” in the **Body content** text box, choose KEYENTITIES from the **Feature** drop down list, and type “hybrid” in the **Feature value** text box.



Body content:

Feature:

Feature value:

The retrieved results are displayed in the [Document Query Result](#) screen.

4 Conceptual Search & Navigation Paradigms

The conceptual search & navigation paradigms (**Structure**, **Patterns**, and **Ontology**) are based on the use of entities from the knowledge base. The knowledge base is a kind of database that stores the knowledge for a particular domain. It consists of a set of data (entities, entity properties, descriptions, aliases) with respect to a conceptual model, called ontology, and rules for reasoning over this data. Ontology is a formal representation of the concepts within a given domain and the relationships between these concepts.

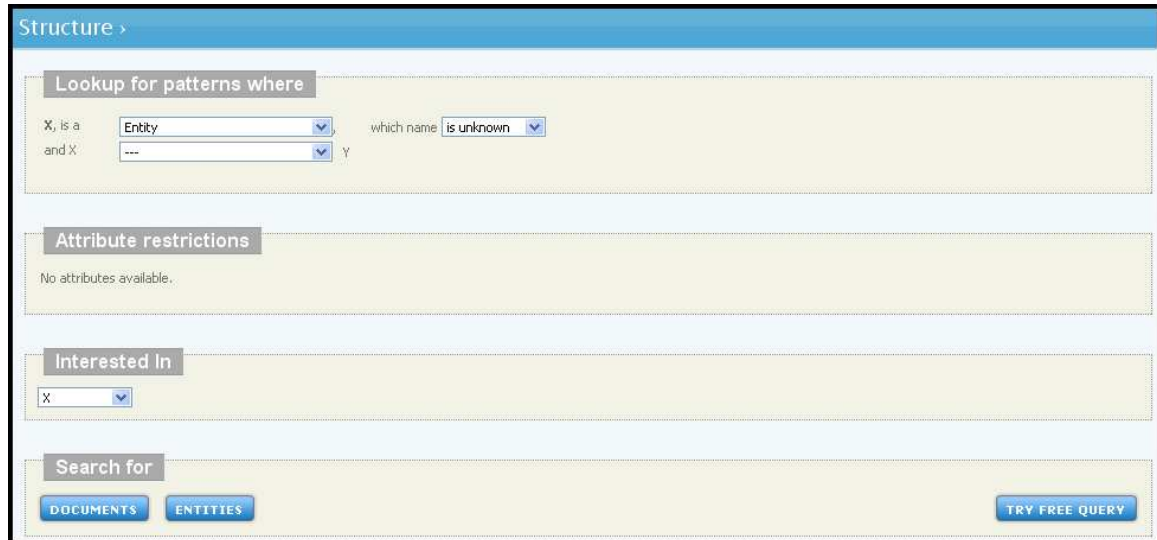
Structure and **Patterns** provide complex tools for retrieving information. Instead of looking for strings of characters in the content of the document set, like the **Keyword** paradigm, conceptual search paradigms search over the graph of the knowledge base and therefore can retrieve entities, as well as documents containing these entities.

Ontology provides a tool for thoroughly exploring the ontology classes of entities and their properties ([PROTON Ontology](#)). The **Ontology** navigation paradigm introduces you to a given domain and helps you create relevant patterns for searching in the data set (especially in the **Structure** paradigm).

4.1 Structure

As seen in the overview, both **Structure** and **Patterns** look for entities in the knowledge base. However, while **Patterns** provides a set of frequently used queries where the search criteria are already pre-defined for-you, **Structure** offers greater flexibility and very comprehensive query definition options. In broad terms, **Structure** can be used for two kinds of search tasks:

- To look for entities through their names, part of their names, or entity type (i.e. entity look-up)
- To create patterns by choosing entity types, names and relations between them. If you want to make your search even more specific, you can use additional entity characteristics.



The **Structure** screen provides three fields where you define your query: **Lookup for patterns where**, **Attribute restrictions**, and **Interested in**.

Lookup for patterns where

The **Lookup for patterns where** field builds the frame of your query pattern. A custom pattern can be represented as:

The entity X is in RELATION with an entity Y, which can be also in another RELATION with another entity Z. Additional restrictions can be specified over X, Y and Z to further narrow the search. This way, each custom pattern is defined by the RELATIONS between the selected entity types and the restrictions over their names.

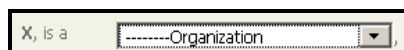
Query patterns for entity X

By default the **Lookup for patterns where** field is set only for the entity X. It consists of three boxes with drop down lists and an optional search box.

the X is a box

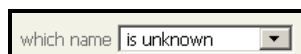
In the **the X, is a** box you specify the entity X by selecting the entity type you are interested in. Here the default is set to "Entity" - the most general ontology class.

If you click the drop down list, you see the taxonomy tree representing the ontology. To make your query pattern more specific, just select one of the classes or subclasses of the ontology, for example "Organization".



which name box

In the **which name** box you further restrict your query by specifying the name of the entity you have selected. This is also a drop down list box and the default is set to: "is unknown".



If you choose this selection, the results will contain either all entities for the type organizations or all documents containing entities for organizations.

However, if you have even some partial information about the name of the entity, you can click the drop down list and select one of the comparison operators:

- *is exactly* (case-sensitive) - if you know the exact name of the entity, including exact upper- / lowercase spelling
- *equals ignore case* - if you know the name of the entity but you are not sure about upper- / lowercase spelling
- *starts with / ends with* - if you know the first/last name (or letters) of the entity
- *contains* - if you know only some part of the name
- *is < (before) / is > (after)* - if you have a time interval

Empty search box

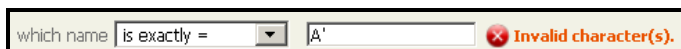
When you select an operator an empty search box opens and you can type the name of the entity there.

TASK: Find specific companies, for example all companies that have Apple in their name.

TO DO: Select “Organization” as your entity type and *contains* as your comparison operator and type “Apple” in the search box.



- When you type the name of the selected entity, you need to enter only strings of letters, numbers, spaces, underscores, and any value (“*”). Otherwise you get an error message.



If you are only interested in the entity X, you can continue defining your query in the **Attribute restrictions** field. However, if your task requires more complex information, you can further restrict the query by adding other entity types and their relations.

Further restricting the query pattern

and X...Y box

To further restrict the entity pattern, go to the **and X...Y** box and from the drop down list select the type of relation between the entity X and another entity Y, applicable to your query.

This opens another section in the **Lookup for patterns where** field. It provides a new set of options for restricting the entity Y like the ones described for the entity X. Here, you specify the name of the entity Y and whether you are interested in its relation to a third entity.

- Alternatively, you can specify the relation of the entity X to the third entity Z by selecting the entity X from the drop down list.

When the last section in the **Lookup for patterns where** field becomes available, you can specify the name of the entity Z in the same way as described for the entity X. This finalizes the frame of your custom query pattern.

TASK: Find all People in the knowledge base that have a Position in specific companies, for example companies with Apple in their name.

TO DO: Enter all restrictions as shown in the following custom query pattern:

- **X, is a Person , which name** is unknown
- **and X** hasPosition **Y**
- **Y, is a Job Position which name** is unknown
- **and Y** withinOrganization **Z**
- **Z, is a Organization which name** contains “Apple”
- **Interested In:** X, Y, and Z
- **Search for** Entities

Attribute restrictions

If there are any attributes available for the entities, you can narrow your custom pattern further in the **Attribute restrictions** field. This field becomes automatically available when you select entities in the **Lookup for patterns where** field.

The section for each entity consists of two boxes with drop down lists and an optional search box. The first drop down list contains the attribute options available in the ontology. The second one consists of the comparison operator and by default is also set to “is unknown”. When you choose another operator, an empty search box becomes available and you can type restrictions on the entity name there.

X	startTime	is < (before)	2009
Y	numberOfEmployees	is > (after)	100

For example, you may be interested in job positions held before 2009 in organizations which name contains “Apple” and where the number of employees is more than 100.

Interested In

In the **Interested in** field you identify which of the entities in your custom query pattern, should be included in the results. You do this by selecting an option from the drop down list.

In the case of the query pattern “Find all People in the knowledge base that have a Position in companies with Apple in their name”, you can retrieve all:

- people (entity X)
- people and job positions (entities X and Y)
- people and organizations (entities X and Z)
- people, job positions and organizations (entities X, Y, and Z)

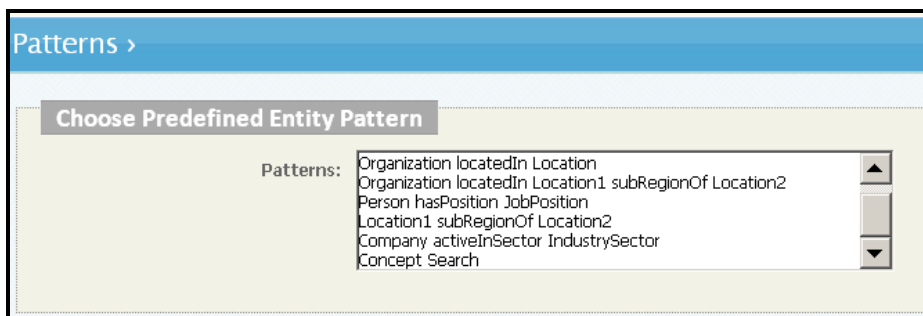
Search for

In the **Search for** field you choose whether to look for entities in the knowledge base or for the set of documents where entities matching your query pattern occur. This field works in the same way in both the **Structure** and the **Patterns** search paradigms.

- To get the set of all documents matching your custom query, click **DOCUMENTS**. Results are presented with some metadata and snippets in [Document Query Result](#).
- To get the set of all entities matching your custom query, click **ENTITIES**. Results are presented with hyperlinks to their entity description in [Entity Query Result](#).
- To further customize your query by entering more restrictions directly in SeRQL, click **TRY FREE QUERY**.

4.2 Patterns

As seen in the overview, both **Structure** and **Patterns** look for entities in the knowledge base. However, while with **Structure** you have to create query patterns of your own by specifying the types of entities, their names and the relations between these entities, **Patterns** provides a set of frequently used queries where the search criteria are already pre-defined for you. These pre-defined query patterns can depend on the domain and how data is represented in the knowledge base. Therefore, they can be customized to meet different needs.



The **Patterns** paradigm screen provides three fields where you define your query. In the **Choose predefined entity pattern** field you select the pattern that suits your search best. Each pattern has different input fields, which appear in the **A pattern about** section after you have selected a pattern. The values in the input fields restrict the types of entities forming the pattern and filter the information. In the **Interested in** field you specify the type of result you want to retrieve.

Choose Predefined Entity Pattern

To choose a pattern from the **Choose predefined entity pattern** field, just click an option from the list of patterns. These patterns are structures of two or three entity types and the relations between them (e.g.: “Organization locatedIn Location” or “Person hasPosition JobPosition withinOrganization Organization”).

A Pattern about

Depending on whether you have selected a structure of two or three entity types, **A pattern about** field consists of two or three empty search boxes respectively. You can restrict the types of entities in the pattern by typing their exact names (labels) in these search boxes.

TASK: Find all organizations active in a particular region, as well as their exact location, for example all organizations active in the Middle East and their exact location.

TO DO: Select the pattern “Organization locatedIn Location1 subRegionOf Location2” and type “*”, “*” and “Middle East” in the respective search boxes as shown below.

A Pattern about

Organization

located in Location1

sub region of Location2

- When you type the name of the selected entity, you need to enter only strings of letters, numbers, spaces, underscores, and any value (“*”).

Interested In

In the **Interested in** field you specify which of the entity types in the pattern you have selected should be included in the results. You do this by selecting an option from the drop down list.

In the example above, you may be interested in:


Patterns

- Organization, Location1, Location2
- Organization, Location1
- Organization, Location2
- Location1, Location2
- Organization
- Location1
- Location2

Search for

In the **Search for** field you choose whether to look for entities in the knowledge base or for the set of documents where entities matching your query pattern occur (as in **Structure**).

4.3 Ontology navigation

The **Ontology** navigation paradigm shows the ontology used in the  showcase. This kind of conceptual modeling represents an abstract, simplified, and structured view of a specific area of knowledge – in this case the news domain. It consists of a set of concepts assumed to exist in it and the relationships between them. As seen in the overview, you can use the **Ontology** paradigm for thorough exploration of the ontology classes of entities and their properties. In this way you can get to know the news domain as well as the conceptual model used to formalize it.



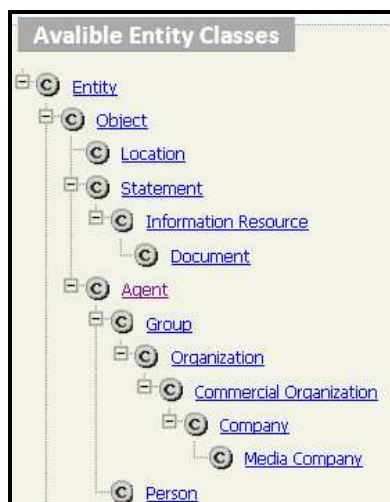
The **Ontology** screen consists of two panels. The first one is the **Available entity classes** field displaying the ontology tree where all classes are presented as hyperlinks. The second one appears when you click one of these hyperlinks and gives you a detailed description of the selected ontology class.

Available Entity Classes

The [PROTON](#) ontology defines about 300 entity classes and 100 properties. All classes that represent categories of objects or phenomena in the applied domain are defined as sub-classes of “Entity”. Then, a further distinction is made between three main entity classes:

- **Object** – existing entities, such as agents, statements, locations
- **Happening** – events, situations, and time intervals
- **Abstract** – abstractions that are neither objects nor happenings

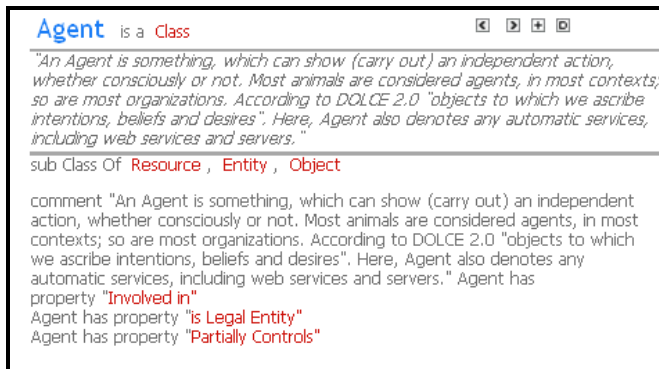
Each of these three is further sub-divided into other classes, representing concepts of general importance such as commercial, government, and other organizations (the class “Organization”), people (the class “Person”), various locations (the class “Location”), different types of employment (the class “Job Position”), etc. There are also different characteristic attributes and relations defined for each class. To learn more about the default ontology, please see [PROTON Guidance](#).




Class description panel

Clicking one of the hyperlinks in the ontology tree opens a new panel with a detailed description of the selected ontology class. The new panel displays the following:





- The main characteristics of the selected class.
- A short definition or what is the accepted meaning of this term in the ontology representation.
- The relations defined for the selected class.



To explore the full list of relations defined for the selected class, click the  icon in the top right corner of the class description panel.

You can also explore all objects that are related to the selected class. The ones you can browse are marked as hyperlinks in red.

You can also use the other icons in the top right corner of the panel in the following way:

- To return to the short list of main relations, use the  icon.
- To see all documents where the entities from the selected class occur, use the  icon. You are directed to the [Document Query Result](#) screen.
- To navigate back and forth through the class descriptions you have already checked, use the   icons.

5 Combined Search Paradigms

Combined search paradigms ([Facets](#) and [Timelines](#)) comprise both the conceptual and the keyword search paradigms. They provide advanced tools for retrieving information about entities, points in time or time intervals, and terms that appear together in the same context in the document set. In other words, Combined search paradigms help you find unknown entities by defining the environment in which they appear (relations with other entities, dates, or terms).

This co-occurrence-based faceted search capability is a method of aggregating a popularity rank for entities. How frequently an entity occurs in a context can signify its association with this context. From the frequency of occurrence you can determine if such relation exists and how strong it is. This way you can track some statistics for the relations between entities and documents.

If you have large amounts of data to work with, you can also analyze some very interesting relations between the entities in the domain that shape trends and dependencies. The scalability of these paradigms allows you to search quickly and effectively for these correlations through extensive sets of documents.

5.1 Facets

As seen in the overview, **Facets** tracks which entities, or entities and terms appear together, when, and how often.

5.1.1 Keyword search

The **Document keyword filter** search box is the place where you enter keywords or document feature values to define your criteria.

When you type your keyword query, this automatically updates the entity columns and shows the top ranking entities for the current selection.

TASK: Find what concepts relate to a particular term and how they co-occur in the news articles, for example what concepts relate to the term “recession”?

TO DO: Type “recession” in the **Document keyword filter** search box.

The following top ranking entries appear in the entity columns: “Ben Bernanke”, “Gordon Brown” and “Alistair Darling” (for People), and “United States” and “UK” (for Locations). It means that the word “recession” co-occurs with these entities in the same context.

The screenshot shows a search interface with a 'Facets' section on the left. The main search results are displayed in a table below the facets. The facets include 'Selected Items', 'People', 'Organizations', 'Locations', and 'Related Concepts'. The 'Organizations' facet is currently active, showing a list of suggestions for 'Fed'.

Date	Title
03-04-2009	Recession a 'relatively mild' challenge for Canada: Flaherty ... 2:05 PM ET CBC News Finance Minister Jim Flaherty said Friday the country is in a mild recession and that Canada will come out of it strongly. Flaherty delivered his upbeat message in ...
04-04-2009	Recession a 'relatively mild' challenge for Canada: Flaherty ... Minister Jim Flaherty, speaking in London, England, on Friday, said this is "a mild economic recession." (Matt Dunham/Associated Press) Finance Minister Jim Flaherty said Friday the country is in a ...
23-11-2008	Share salvation: returns that can defy recession Share salvation: returns that can defy recession From gold to fags and booze to home entertainment, some investments can make you ...

To clear your query, click the **Clear** link in the right hand corner of the screen or just press DELETE.

5.1.2 Co-occurrence of entities

The columns (e.g. People, Organizations, etc.) represent the top ranking entities of particular types in the set of documents. The entity types are classes in the domain ontology. When you enter parts of entity names or when you select an entity or type a keyword, the facets are refreshed.

When you start typing an entity in the empty search box of each column, you get an auto-complete list of suggestions with the typed part of the entity name in highlight.

TASK: Find information about a specific organization, for example the “Federal Reserve”.

TO DO: Start typing “fed..” in the search box of the **Organizations** column and select the relevant entry for the list of suggestions such as “Federal Reserve”, “Federal Deposit Insurance”, “Federal Bureau of Investigation”, etc.

The close-up shows the 'Organizations' facet search box with the text 'Fed' entered. Below the search box, a list of suggestions is displayed, including 'The Fed', 'Federal Reserve', 'Federal Deposit Insurance', 'Federal Bureau of Investi...', 'Federal Open Market Commi...', 'Federal Trade Commission', and 'FedEx'.

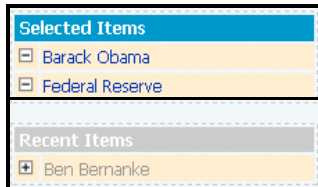
When you select an entity, it appears in the **Selected items** list. This limits the list of entities in each column to only the entities occurring together with it. The **Related concepts** column shows updated information on concepts semantically related to the selected entity. This column is not configurable.

Finally, the document count is refreshed showing only the documents in which this entry appears.

TASK: Find information about a specific organization and a political figure, for example the Federal Reserve and Barack Obama.

TO DO: Type “Barack Obama”, “Federal Reserve” in the search box of the **Persons** and **Organizations** entity class columns respectively.

To change the restrictions of your query and remove an item from the **Selected items** list, just use the icon and the item is added to the **Recent items** history.



To add it back to the **Selected items** list, just click the icon next to the entry.

The auto-complete function will not offer you a list of suggestions for numbers as it does for letters. If the name of the organization you are looking for starts with a number as in, for example “454 Corporation”, when you start typing “454” in the search box of the **Organizations** column, there will be no matches found. The same will occur if you want to search in the **Happening** column for an event by date.

5.1.3 Co-occurrence of entities and keywords

If you want to use the full potential of **Facets**, make your search criteria more complex by creating combinations of keywords and entities. To define your search, just type or select entities from the columns and type keywords in the **Document keyword filter** box.

TASK: Find all documents that have a specific term in their title and mention location and person in their body, for example the term “recession” to co-appear with Asia and Barack Obama.

TO DO: Select “Asia” and “Barack Obama” from the respective columns and type “TITLE:recession” in the **Document keyword filter** box.

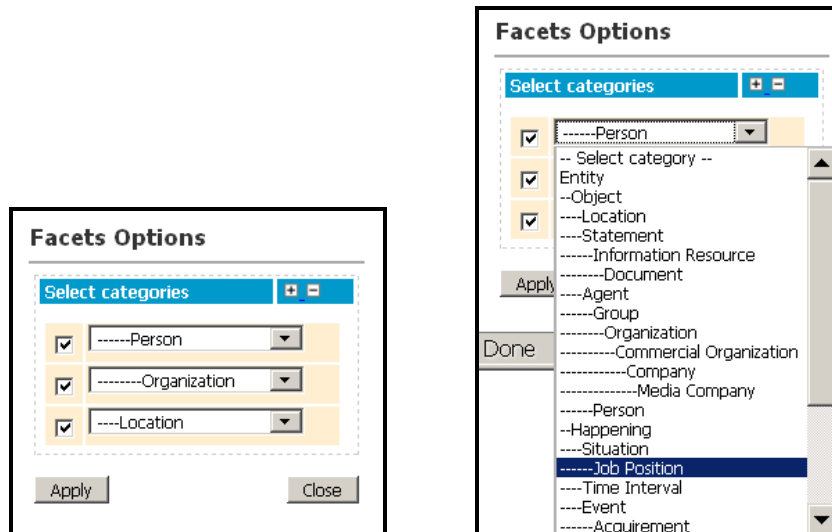
The screenshot shows the search interface with the following components:


- Selected Items:** Barack Obama, Asia
- Recent Items:** (No recent items)
- People:** 14 shown below, including Barack Obama, Timothy Gathner, Ken Hensley, Christina Romer, David Weiss, Allen Sinai, Martin Orlinger, Ben Bernanke, Oliver Blanchard, Herb Goldschlus, Kevin Rudd, Andrew Livens, George W. Bush, Brian Behlune.
- Organizations:** 24 shown below, including The Associated Press, Congress, Global Technologies, Ltd., Department of Treasury, European Union, Monetary Fund, Dow Jones & Co., Inc., Outlook Group Corp., World Bank, The Fed, Business Breaking News, ClearView Economics, European Central Bank, General Motors, Chrysler Corp., Department of Labor, White House Council Of Ec..., Dow Chemical Co., National Association, United States Senate, Department of Commerce.
- Locations:** 18 shown below, including Washington, United States, Europe, Canada, United Kingdom of Great B..., Federal Republic of Germa..., People's Republic of Chi..., Republic of India, Ciudad de Mexico, White House, Japan, North America, Mexico, Frankfurt, Wall Street, Departamento de Florida.
- Related Concepts:** Barack Obama, Washington, Percent, United States, Asia, Recession, Next Year, The Associated Press, Congress, Economy, Percent, \$927 Billion, 2009, Department of Treasury, 2.5 Percent, Second Quarter, Timothy Gathner, Europe, 2010, The 1990s, Global Technologies, Ltd., 25 Million, Republic of India, 8.9 Percent, One, Japan.
- Document Keyword Filter:** TITLE:recession
- Documents, containing all selected entities:** 1.5 of 5 documents matching the search criteria.
- Matching documents:** 5. Between: 23 Apr, 2008 - 02 Jul, 2010.
- Search Results Table:**

Date	Title
22-04-2009	Global recession worst since Depression, IMF says WASHINGTON (AP) -- The global economy is expected to ...
23-04-2009	Global recession worst since Depression, IMF says WASHINGTON (AP) -- The global economy is expected to ...
15-11-2008	Europe officially in recession /W3C/DTD XHTML 1.0 Transitional/EN" http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd"> Europe officially in recession Business Breaking News News.com.au Welcome to NEWS.com.au Business. Skip to: Search Section ...
29-10-2009	Economy grows in Q3, signals end of recession Economy grows in Q3, signals end of recession WASHINGTON (AP) -- The economy grew at a 3.5 percent pace in the third ...
30-10-2009	The recession may be over at last... so what now? The recession may be over at last... so what now? WASHINGTON (AP) -- After a ...

5.1.4 Configuration options

The columns of entity classes are optionally configurable. To specify the entity class for each column, click the **Options** link in the top right corner of the **Facets** screen. This opens a new **Options** screen that shows the default columns. To select new categories, click the arrow next to each column heading and make your selection from the drop down list representing the [ontology](#) classes. Click **Apply** and then **Close**.



To add more columns, click the  icon and a new field appears. Select a category for it and click **Apply**. Now you have four columns in the **Facets** screen.

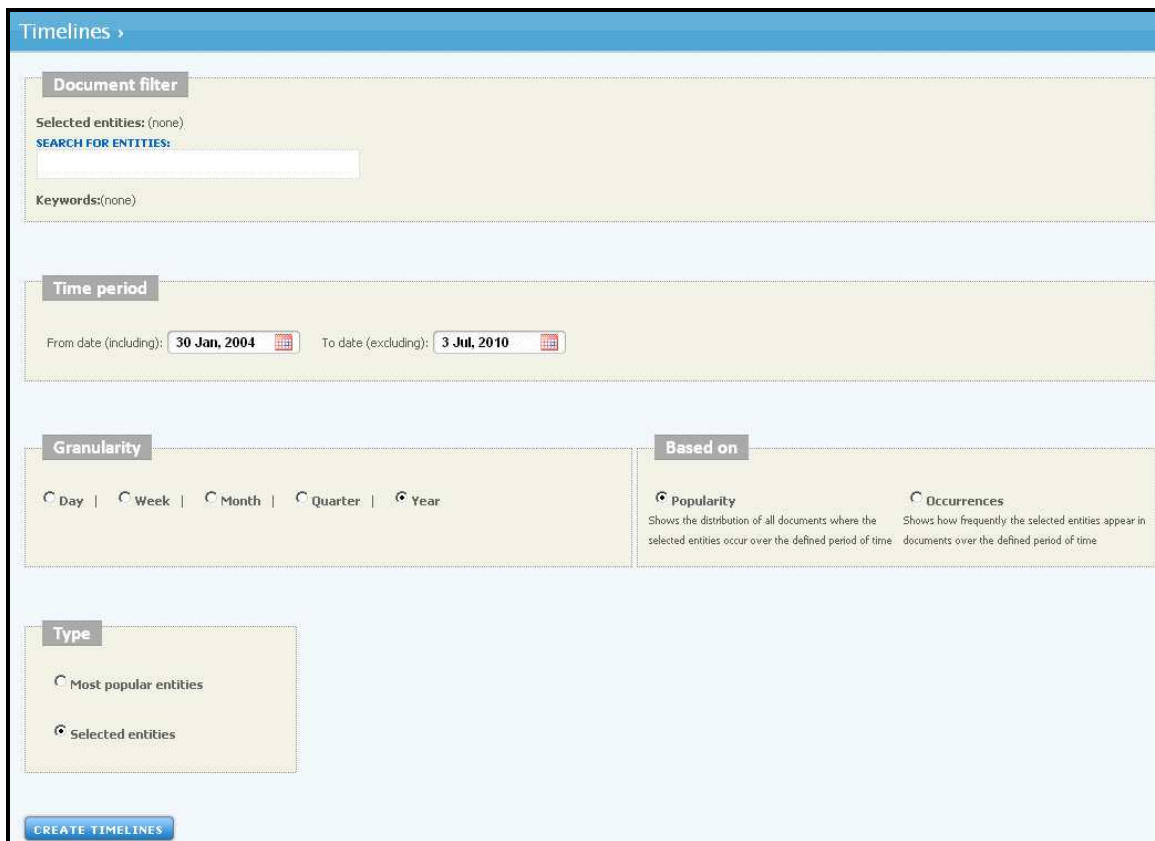
To the default number of columns, click the  icon and you have three columns again.

To view the results of your query, go to the [Document Query Result](#) or the [Timelines Result](#) sections.

5.2 Timelines

Timelines is a tool for analyzing statistics for entities found in texts over a period of time. This statistics is based on how frequently entities appear together in the document set. In general, documents are bound to concrete moments in time. The **Timelines** feature uses this additional information (metadata) about the date when each document was created. Thus it analyzes how documents and entities are distributed over a particular period and what types of co-relations exist between them.

The **Timelines** components can be formed either by pre-selected entities or by the most popular entities in a given document set and time frame. For each period one can further examine the underlying set of documents forming the part of the timeline.



The screenshot shows the 'Timelines' interface with the following sections:

- Document filter:** Includes 'Selected entities: (none)', a 'SEARCH FOR ENTITIES:' input field, and 'Keywords:(none)'.
- Time period:** Shows 'From date (including): 30 Jan, 2004' and 'To date (excluding): 3 Jul, 2010'.
- Granularity:** Radio buttons for Day, Week, Month, Quarter, and Year (selected).
- Based on:** Radio buttons for Popularity (selected) and Occurrences. Descriptions: 'Shows the distribution of all documents where the selected entities occur over the defined period of time' and 'Shows how frequently the selected entities appear in documents over the defined period of time'.
- Type:** Radio buttons for Most popular entities and Selected entities (selected).

A 'CREATE TIMELINES' button is located at the bottom left.

Generally, you can calculate three types of timeline analysis:

- **Documents timelines** calculates how the documents, in which the selected entities occur, are distributed over a particular period of time.
- **Selected entities timelines** calculates trends for selected entities over a particular period of time.
- **Most popular entities timelines** calculates trends for the most frequently mentioned entities of a given type over a particular period of time.

5.2.1 Creating a timeline query

In order to generate a timeline query, you have to:

- Define the entity or entities you want to look for by entering them in the **Document filter** field.
- Specify the time period over which you want to perform the query by defining the start and the end date in the **Time period** field.
- Define your search criteria further by setting your options in the **Based on** and **Type** fields.
- Configure the time units into which the time period is divided in the generated chart by using the **Granularity** options.

Document filter


The **Document filter** field consists of a search box and a text field displaying details about the current filter of the query.

You enter the entity you will be looking for in the **Search for entities** box. As you start typing, the auto-complete feature automatically recognizes possible names of entities and offers a list of suggestions. Selecting an entity from the list, adds it to the **Document filter**.


TASK: Find how the popularity of a political figure changes through time in the international news, for example Barack Obama.

TO DO: Start typing “oba” and then select “Barack Obama” from the list of suggestions.




 You can use an unlimited number of entities in a query by adding them to the Selected entities in the **Document filter**.





To remove any of the selected entities, click the  icon next to it.

Time period

The **Time period** field consists of two search boxes: **From date** and **To date**. They set the start and the end of the time period over which your query is performed.

To define the start point of the time period, go to the **From date** box and click the  (calendar) icon. This opens a drop down menu with a calendar view of the currently selected month and year. On top of the calendar view there are two drop-down list boxes. The first one selects the month and the second – the year.



 The set of possible choices for years is defined by the corpus of all documents annotated by  latestnews. Just have in mind that when setting the time period, you have to define the whole date (month, year, and day), even if you want to select just a different year. If you change only the year without setting a day and a month, when you close the calendar icon, your year selection will not be taken into account.

TASK: Find how the popularity of a political figure and an organization changes through time, for example Barack Obama and the Federal Reserve in the time period between 1 January 1st, 2009 and June 30th, 2010.

TO DO: Click the **From date** search box and select “January” from the list of months, “2009” – from the list of years, and 1st from the calendar menu respectively. After that, set the end date (June 30th, 2010) in the same way in the **To date** search box.

Granularity

The **Granularity** field gives you the option to view the set of timelines results broken into smaller time units such as days, weeks, months, quarters, or years.

- Days – when the documents, in which the selected entities occur, cover a time period of less than 40 days
- Weeks - when the documents, in which the selected entities occur, cover a time period of more than 40 days and less than 6 months
- Months - when the documents, in which the selected entities occur, cover a time period of more than 6 months and less than 2 years
- Quarters - when the documents, in which the selected entities occur, cover a time period of more than 2 and less than 5 years
- Year - when the documents, in which the selected entities occur, cover a time period longer than 5 years

Based on

The **Based on** field provides two options – popularity and occurrence. The first one shows the distribution of all documents where the selected entities occur over the defined period of time, while the second option - how frequently they appear.

For example, when you want to find how the popularity of Barack Obama and the Federal Reserve changes between 1 January 1st, 2009 and June 30th, you can view timelines for all documents in which these two entities occur within this time period. Alternatively, you can analyze the frequency of appearance of these two entities in documents in this time frame.

Type

The **Type** field also offers two options. The first one is to perform your analysis based on the entities you have specified in the **Document filter** field. The second option is to analyze the most popular entities of an entity type over the defined period of time. When you select **Most popular entities**, a new field – **Options** – appears and enables you to make further restrictions over your query.

TASK: Analyze the trends and interrelations between the most frequently mentioned entities of the type “Person” in the news over the period between 1 December, 2008 and 31 March, 2009.

TO DO: Set your time period and granularity, and choose whether the results are based on popularity or occurrence. Then select **Most popular entities** in the **Type** field.



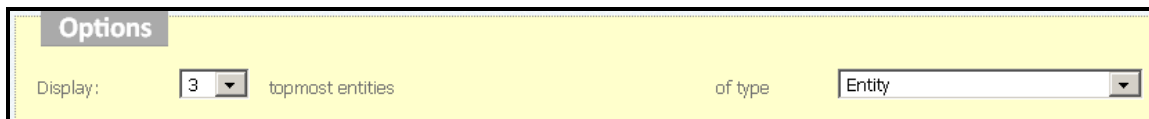
The screenshot shows a panel titled "Type" with two radio button options: "Most popular entities" (which is selected) and "Selected entities".

The **Timelines** analysis of most popular entity calculates ups and downs in the document flow and gives a relatively realistic evaluation of the changes in the popularity of an entity.

Options

When you select the **Most popular entities** option from the **Type** field, you can define further restrictions over your query in the **Options** field that opens:

- **Display...topmost entities** – specifies how many of the top ranking entities are displayed at the same time. You can select a number between 1 and 20 from a list box.
- **Of type** –you can choose the type of entity for which popularity over a particular time frame is analyzed. You can select the type from a drop down list representing the ontology classes.




The screenshot shows an "Options" panel with two fields: "Display:" with a dropdown menu set to "3" and the text "topmost entities" next to it, and "of type" with a dropdown menu set to "Entity".

5.2.2 Timeline results

You can view the results in the **Timelines Results** screen, presented in a chart. The chart shows the fluctuation of the entities’ frequency of appearance under the given restrictions.

From there you can also navigate to the documents that formed the respective peaks and drops in popularity.

6 Search Results

 latestnews provides three kinds of search results, depending on the type of search you perform:

- **Document Results:** show the retrieved documents matching a query.
- **Entity Results:** show all the entities of a given class.
- **Timelines:** provide a graphical representation of how entity popularity changes through time in a given set of documents.

6.1 Document Query Result

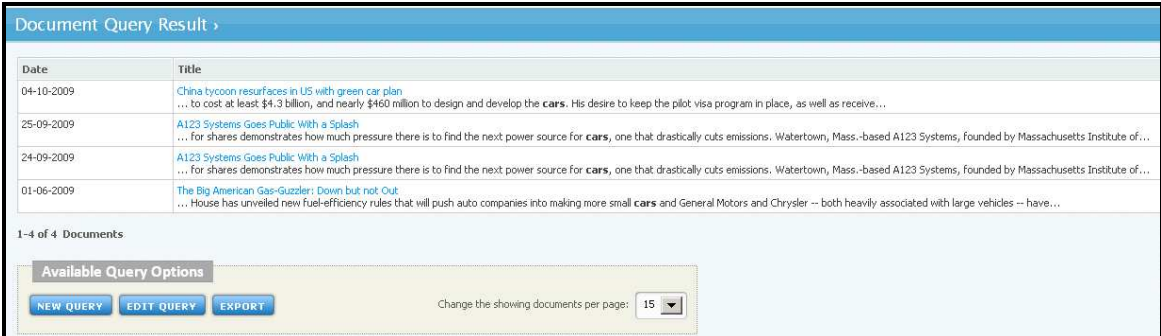
Document Query Result presents the documents corresponding to your search criteria and is the most common type of search result.

- The examples in this section are based on a **Boolean** search for “cars” with “KEYENTITIES:hybrid”.

Table of results

The document results are arranged a **Table of Results** featuring the following columns:

- **Date column** - shows the date when the document was created.
- **Title column** - shows the title of the retrieved documents and snippets from the document content, displaying where the selected entities/keywords are mentioned. The title is marked as a hyperlink to the document. If you click the link, you go to the **Document Detail** screen.



Date	Title
04-10-2009	China tycoon resurfaces in US with green car plan ... to cost at least \$4.3 billion, and nearly \$460 million to design and develop the cars. His desire to keep the pilot visa program in place, as well as receive...
25-09-2009	A123 Systems Goes Public With a Splash ... for shares demonstrates how much pressure there is to find the next power source for cars, one that drastically cuts emissions. Watertown, Mass.-based A123 Systems, founded by Massachusetts Institute of...
24-09-2009	A123 Systems Goes Public With a Splash ... for shares demonstrates how much pressure there is to find the next power source for cars, one that drastically cuts emissions. Watertown, Mass.-based A123 Systems, founded by Massachusetts Institute of...
01-06-2009	The Big American Gas-Guzzler: Down but not Out ... House has unveiled new fuel-efficiency rules that will push auto companies into making more small cars and General Motors and Chrysler -- both heavily associated with large vehicles -- have...

1-4 of 4 Documents

Available Query Options

NEW QUERY EDIT QUERY EXPORT

Change the showing documents per page: 15

Document distribution

When you go to the **Document Query Result** screen from **Timelines Results** by clicking a data point in the chart, you view a graphic representation of the distribution of documents matching your query over a period of time.

The vertical axis shows the number of documents per time period and the horizontal axis - the range of the time period in which these documents occur.

The example below shows the number of retrieved documents for “Barack Obama”, covering the pre-defined time period “from September, 2008 to March, 2009” and the granularity is set to: “quarter”.




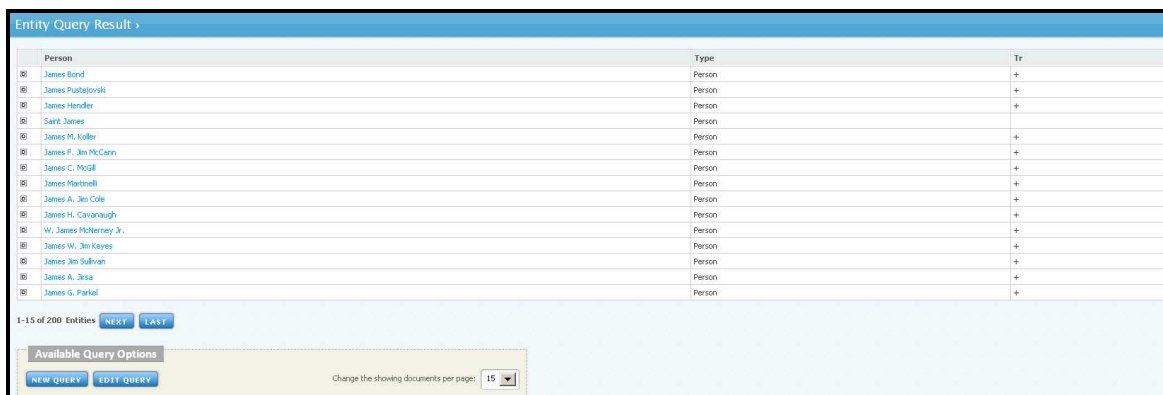
To see the distribution of these documents in further detail, click the month you are interested in and the granularity changes. The new diagram shows the document distribution per day for the month you have selected.

6.2 Entity Query Result

Entity Query Result presents the entities of a given class (type) and it is common for **Structure** and **Patterns**.

The entity results are arranged in a table. If you have selected only one entity in the **Interested in** field from the **Structure/Patterns** screen, the table of results features the following columns:

- The **Documents** column is marked by the icon , which is a link to a list of all documents, where the selected entity or its aliases appear together.
- The **Selected entity type** column shows the names of all entities from the selected entity type in each row of the table. The entity name is marked as a hyperlink and when you click it, it opens an Explorer screen with the **Entity Description**.
- The **Type** column shows the most specific class to which the entity belongs.
- The **Tr** column shows if the entity was pre-populated in the knowledge base from a trusted source (marked by a '+' icon) or was automatically extracted from the documents, processed by the system (without a '+' icon). Due to natural limitations, the automatically extracted entities may represent an incorrect or imprecise model of the world.




Person	Type	Tr
James Bond	Person	+
James Pusteljovski	Person	+
James Hendler	Person	+
Saint James	Person	
James M. Koller	Person	+
James F. Jim McCann	Person	+
James C. McGill	Person	+
James Martelli	Person	+
James A. Jim Cole	Person	+
James H. Cavanaugh	Person	+
W. James McInerney Jr.	Person	+
James W. Jim Hayes	Person	+
James Jim Sullivan	Person	+
James A. Jim	Person	+
James G. Parkal	Person	+

1-15 of 200 Entities [NEXT](#) [LAST](#)

Available Query Options

[NEW QUERY](#) [EDIT QUERY](#) Change the showing documents per page: 15

If you have selected more than one entity in the **Interested in** field from the **Structure/Patterns** screen, the table of results features additional three columns (**Selected entity type**, **Type**, and **Tr.**) for each of the selected entities.

By default  **latestnews** restricts the number of results matching the query to a maximum of 200. This parameter can be changed by the administrator.

6.3 Timelines Result

Timelines Result provides a graphical representation of how entity popularity changes through time in a given set of documents. These results can be generated from a query created in [Facets](#) or directly in [Timelines](#).

The set of results in **Timelines Result** is presented in a chart. Each data point in the chart corresponds to a snapshot of documents, matching the query in a particular moment in time. When a data point is selected, it shows the list of available documents or entities for this moment in time.

The **Timelines Result** screen consists of a chart, representing popularity trends for the selected documents or entities and a text field displaying the document filter and the time criteria defined in the query.

If you want to adjust the default criteria to fit your task better, go to the link “click here” and it takes you to [Timelines](#).

Alternatively, you can generate **Timelines** analysis by selecting **Timelines** directly from the main menu tabs and defining your query and timeline criteria there.

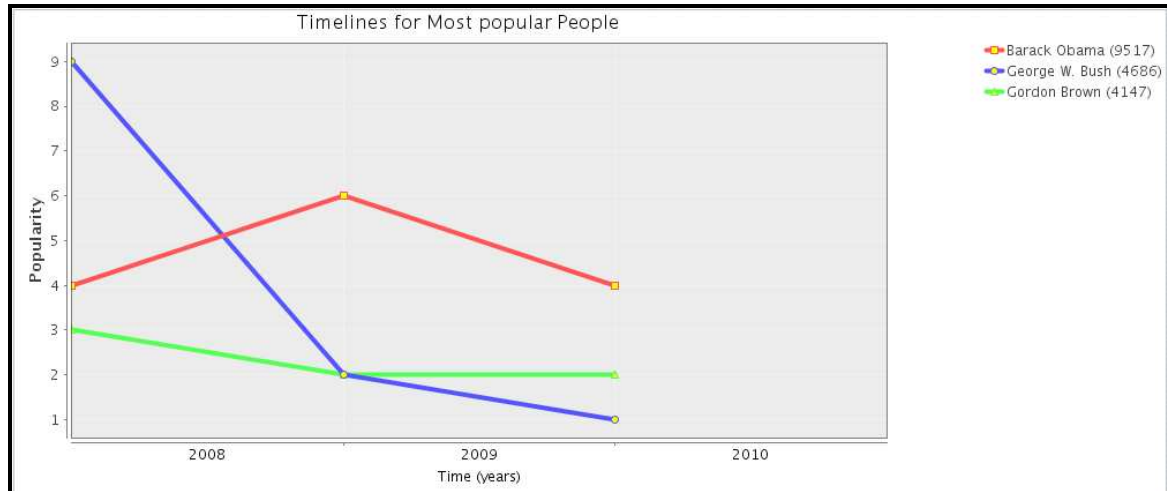
Timelines chart

The chart represents fluctuations in the popularity of selected entities and the documents where they appear under the given restrictions. Depending on whether the **Timelines** analysis is based on popularity or occurrence, the lines illustrate:

- how documents, in which each of the selected entities is mentioned, are distributed over the specified period of time. **Popularity** measures the number of documents in percentage and **Time** shows the granularity you have selected for the period.
- how frequently different entities appear in the text. **Occurrence** measures the number of times (hits) each entity is mentioned and **Time** shows the granularity you have selected for the period.

Each data point in the chart shows a snapshot of documents (in percentage) or entities (in number of hits) matching the query in a particular moment in time (on a particular date).

The timeline chart below shows the 3 top most popular people in the international news between March, 2008 and July, 2010. These results are displayed in granularity of year. At the top right of the screen, you can see a legend, showing the selected entities and the color code for each of them.



When the time period values of two entities to coincide, only one of the timelines will be visible as other one overlaps it.

Timelines criteria

The text field under the chart displays the timeline query criteria.

For entities, selected with **Facets** or **Timelines** search:

- **Time period** - shows the start and the end of the time period over which the query is performed.
- **Total** – shows the total of documents for the restrictions period.
- **Granularity** – shows the selected time unit into which the retrieved results are viewed, such as days, weeks, months, quarters, or years.

Time Period:	14/03/2008 to 03/07/2010
Total:	156355 docs for the restrictions period
Granularity:	Year
Options:	display 3 topmost entities of type People for the whole period

For the most popular entities, selected with **Timelines**, the text field has one more section:

- **Options** – shows the number of displayed entities, the type of entities, and the scope of viewing the results.


Browsing timelines results

When you move the mouse pointer to a data point in the chart, you get a tooltip displaying the following information:

- For **Timelines** queries based on popularity, the tooltip shows the name of the selected entity, the particular moment in time (the date), and the percentage of documents (in relation to the whole time period) in which the selected entity occurs on that date.
- For **Timelines** queries based on occurrence, the tooltip shows the name of the selected entity, the particular moment in time (the date), and the number of times (hits) the selected entity appears in different documents on that date.

If you click a particular data point, you go to the [Document Query Result](#) screen.

7 Results Detail

Based on the two basic types of search: keyword and conceptual,  latestnews also offers you the possibility to further navigate in your searched results. You can explore:

- the retrieved document - as metadata and content
- the entities – as graphs in the knowledge base (aliases, linking to other entities, URI, further navigation to documents, etc.)

7.1 Document Detail

The **Document Detail** screen provides a detailed view of the document you have selected from the retrieved document results. It consists of two fields. The document feature list shows additional information (metadata) about the document and the **Body** field - the text of the document.

The example below displays the result of a **Facets** search for Barack Obama and Hamas.

Document features field

Date	22-09-2009
Title	Mideast summit unlikely to relaunch peace talks
Source	Associated Press
Language	english
Author(s)	-- NO AUTHORS --
URL	http://hosted.ap.org/dynamic/stories/U/US_MIDEAST_US?SITE=NYWNE&SECTION=HOME&TEMPLATE=DEFAULT (expired)
Key Phrases	peace talk, meeting, trilateral meeting, resumption, talk
Key Entities	Shimon Peres, Sani Abu, Yasser Abed Raddo, George Mitchell, Benjamin Netanyahu, World Bank, Monetary Fund, Israel Army Radio, Israel, West Bank, U.S., NEW YORK, Washington, New York, Jerusalem

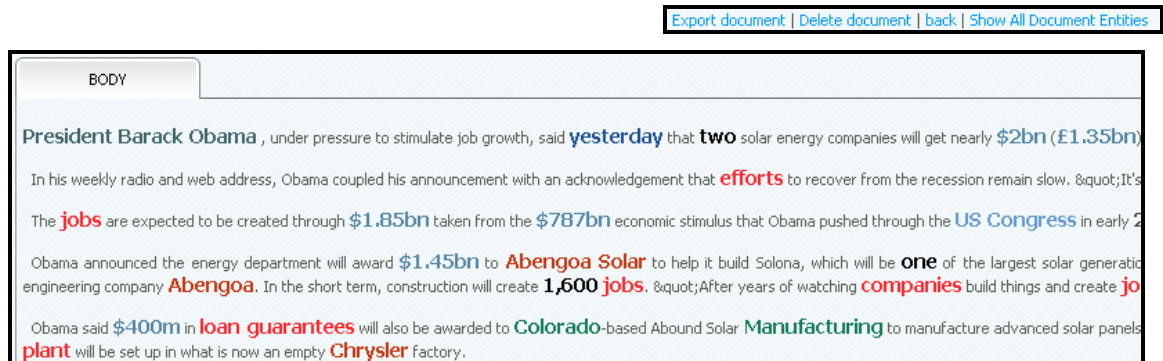
The document feature list provides the following metadata about the selected document: date, title, source (news agency), language, author, URL, key phrases (words or phrases considered characteristic of the selected document), and key entities (named entities that are statistically rare in the document set and therefore are considered to be particularly characteristic of the selected document).

Body field

The **Body** field displays the content of the retrieved document. All entities that were specified in the search criteria are marked as hyperlinks. If you follow these hyperlinks, you can see the entity description of the respective entities. In the example below, the entity “Barack Obama” appears as a hyperlink.

BODY
<p>President Barack Obama, under pressure to stimulate job growth, said yesterday that two solar energy companies will get nearly \$2bn (£1.35bn).</p> <p>In his weekly radio and web address, Obama coupled his announcement with an acknowledgement that efforts to recover from the recession remain slow. &quot;</p> <p>The jobs are expected to be created through \$1.85bn taken from the \$787bn economic stimulus that Obama pushed through the US Congress in early 2009.</p> <p>Obama announced the energy department will award \$1.45bn to Abengoa Solar to help it build Solona, which will be one of the largest solar generation plants in the world. &quot;</p> <p>short term, construction will create 1,600 jobs. &quot;After years of watching companies build things and create jobs overseas, it's good news that we've at</p>

To view all entities recognized in the document content, click **Show all document entities** in the top right corner of the **Document Detail** screen. This shows all entities recognized in the text with hyperlinks to their entity description.



To return to the default view, you can click **Show searched entities only** and you can see again only the selected entities in the content of the document.

7.2 Entity Description

Entity Description shows the **Ontology** class to which the selected entity belongs to, the main relations defined for it, and a list of related entities. You can view it from:

- **Facets** by clicking the red magnifying glass in front of each entity name in the columns.
- **Entity Query Result** by clicking an entity marked as hyperlink in blue in the **Entity** column.
- **Document Detail** by clicking an entity marked as a hyperlink in blue in the **Body** section.

Entity Description consists of two sections. The first section displays the preferred synonym, the definition of the concept, and a list of the main relations defined for the selected entity. The second section shows the entities to which the selected entity is related.

United States is a [Country](#) , [Trusted](#)^{fp}

type [Resource](#) , [Entity](#) , [Object](#) , [Location](#) , [Country](#) , [Political Region](#)

label "United States"

has Alias [United States](#) , [US](#) , [U.S.](#) , [U.S.](#) , [U.S.](#) , [USA](#) , [U.S.A.](#) , [U.S.A.](#) , [America](#) , [United States of America](#)

has Main Alias [United States](#)

Generated by [EnhGazetteer](#)

Part of [North America](#)

Located in [North America](#)

Subregion of [North America](#)

has Capital [Washington, D.C.](#)

has Adjective [Americano](#) , [Americain](#) , [Americains](#)

URI: http://www.ontotext.com/kim/2006/05/wkb#Country_T.4

Related Entities

- [Alabama](#) Part of [United States](#)
- [Alaska](#) Part of [United States](#)
- [Arizona](#) Part of [United States](#)
- [Arkansas](#) Part of [United States](#)
- [California](#) Part of [United States](#)
- [Colorado](#) Part of [United States](#)
- [Connecticut](#) Part of [United States](#)
- [Delaware](#) Part of [United States](#)
- [Florida](#) Part of [United States](#)
- [Georgia](#) Part of [United States](#)
- [Hawaii](#) Part of [United States](#)
- [Idaho](#) Part of [United States](#)
- [Illinois](#) Part of [United States](#)
- [Indiana](#) Part of [United States](#)
- [Iowa](#) Part of [United States](#)
- [Kansas](#) Part of [United States](#)
- [Kentucky](#) Part of [United States](#)
- [Louisiana](#) Part of [United States](#)
- [Maine](#) Part of [United States](#)
- [Maryland](#) Part of [United States](#)
- [Massachusetts](#) Part of [United States](#)
- [Michigan](#) Part of [United States](#)
- [Minnesota](#) Part of [United States](#)
- [Mississippi](#) Part of [United States](#)
- [Missouri](#) Part of [United States](#)
- [Montana](#) Part of [United States](#)

You are seeing the first 52 resources related to the entity! [Click here](#) to see the full list.

Now you are ready to explore the KIM functionalities through the Web UI on your own.

Enjoy!

The Kim Team @ Ontotext