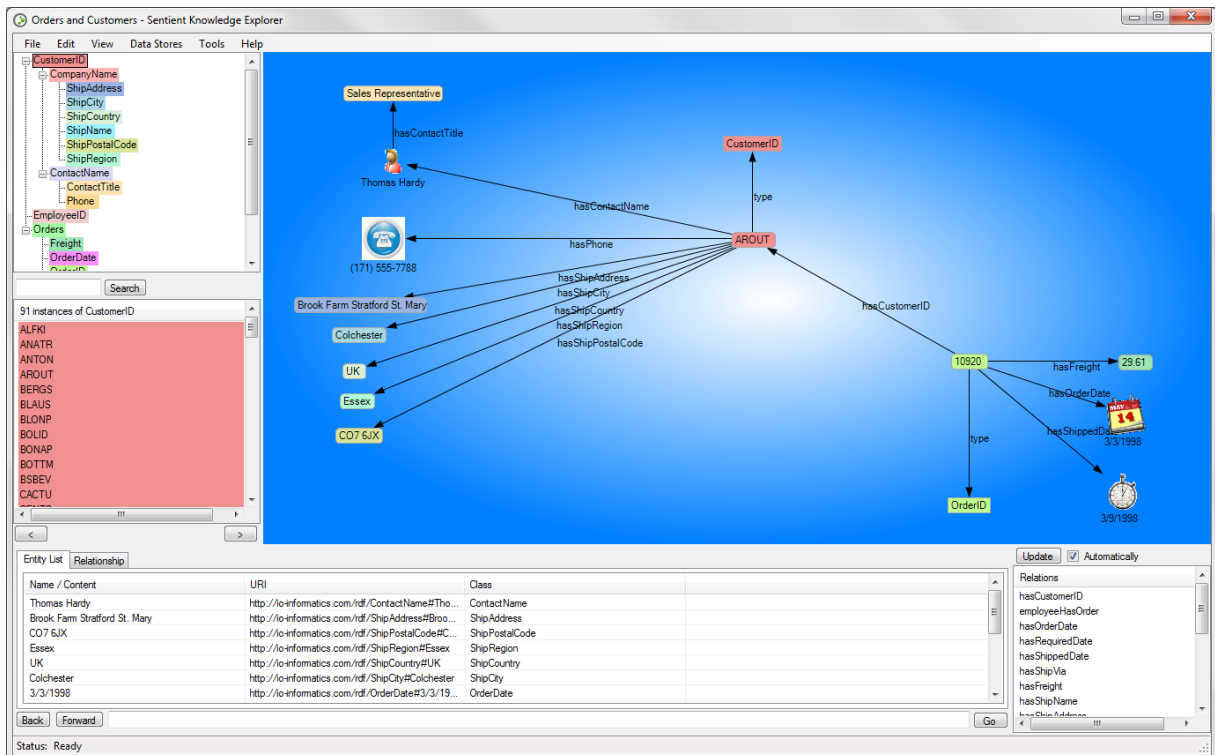


SENTIENT KNOWLEDGE EXPLORER

USER TUTORIAL

The Sentient Knowledge Explorer is a tool designed to let informaticists and other IT experts easily create complex data integrations and applications using the power and flexibility of Semantic Web technologies. In this tutorial you'll see how easy it is to integrate two data sets (customer and order information) to create a Knowledge Base and visualization similar to the image below. These same techniques can be used to create both personal and enterprise-scale integrations and applications both quickly and at a low cost.



Customer Support:
 For the Personal Edition, go to
forums.io-informatics.com
 For the Professional Edition, please use
 the forums or contact
 IO Informatics directly at
 E-Mail: support@io-informatics.com
 Phone: +1 510 705 8470

IO Informatics, Inc.
 2550 Ninth Street, Suite 114
 Berkeley, CA 94710-2552, U.S.A.
www.io-informatics.com

Table of Contents

Introducing the Knowledge Explorer	2
Importing and Mapping data from Spreadsheets	3
The Import Mapper Dialog Window	4
Customizing Your Display with Icons	9
Saving Datasets	11
Working with existing semantic data files	13
SPARQL queries.....	14
Next Steps.....	16

Introducing the Knowledge Explorer

We hope you enjoy using the Sentient Knowledge Explorer. It's a visual tool that allows you to easily integrate multiple data sources in a using a flexible, standard data model. This data model, based on *Semantic Web* or *Web 3.0* technologies, allows you to continually extend your integrations, which we call Knowledge Bases, as well as easily share this data with others and link to data shared on the web. This is an agile alternative to creating data warehouses with relational technology. Due to the flexible nature of RDF and OWL, Knowledge Bases can easily be extended as requirements change without complex refactoring projects.

You will be able to follow the steps in this tutorial without a background in the Semantic Web, but to fully take advantage of the Knowledge Explorer, users will need an understanding of the underlying Semantic Web data model. IO Informatics offers training in these standards and how to most effectively apply them. You can also start to learn about the Semantic Web at:

<http://www.w3.org/standards/semanticweb>

A basic understanding of RDF is helpful for getting the most out of this tutorial. For an introduction to the Linked Data community, go to:

<http://linkeddata.org>

Once you have used the Knowledge Explorer to build a Knowledge Base around your area of interested you can:

- Create rich pattern-based queries to mine your data.
- Look for novel links in your data and form hypotheses.
- Query out to other sources to enrich that data.
- Share the data with colleagues and publish that data on the web using IO Informatics' other tools.
- Build your own custom applications on top of your new Knowledge base.
- Very easily create attractive, publishable views of your data.

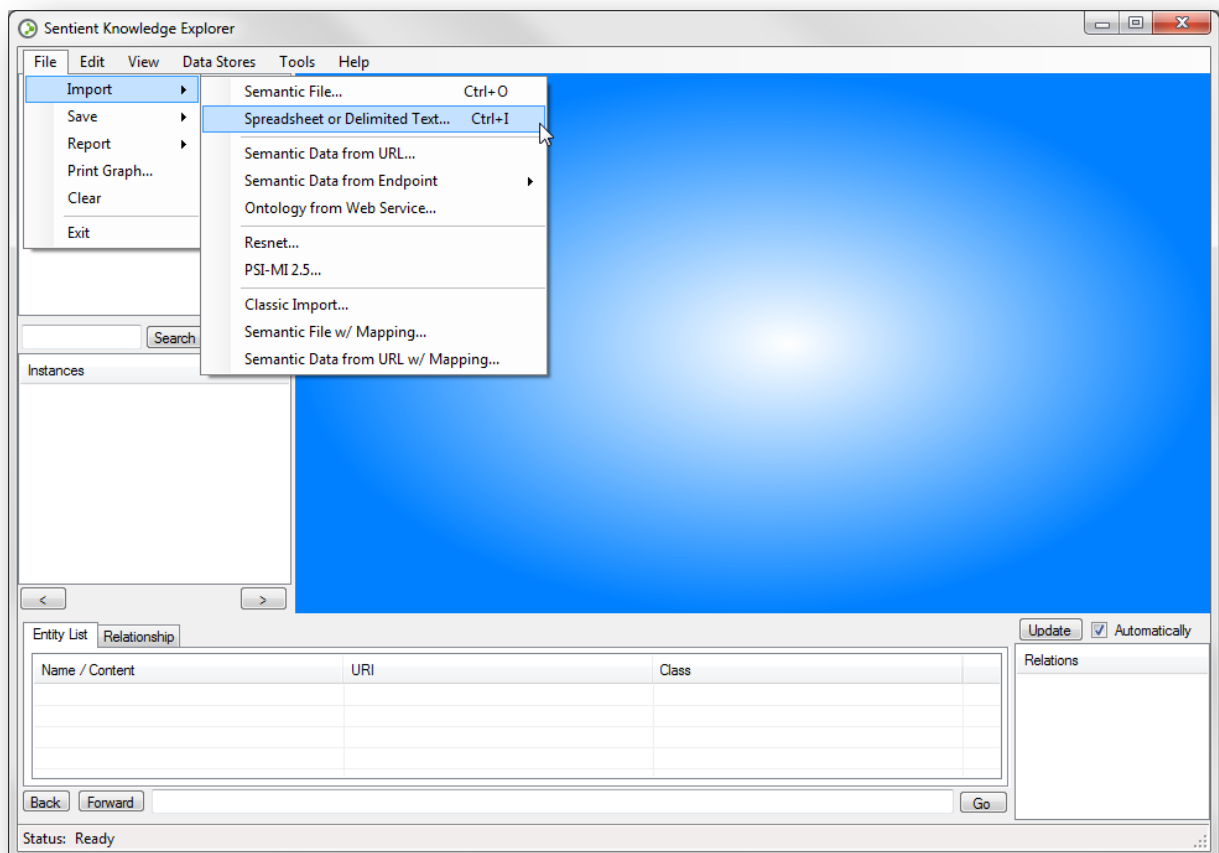
The rest of this tutorial will introduce you to the process by enumerating the steps to create a simple Knowledge Base from two sample files, showing you how easy it really can be to transform multiple data sources into a unified data set. Specifically, you'll integrate customer information with order information from an imaginary business (shown on the title page), then view and query the new Knowledge Base.

Importing and Mapping data from Spreadsheets

Building semantic data sets from extant files is an easy yet powerful tool to assemble linked data. The *Sentient Knowledge Explorer* makes it easy to import data from Excel documents or delimited text files. To demonstrate this, example file comes with each installation of the Knowledge Explorer. With a typical installation this file is located along the path:

\Program Files\IO Informatics\Sentient Knowledge Explorer\Resources\Sample
Data\NW_Customer.txt

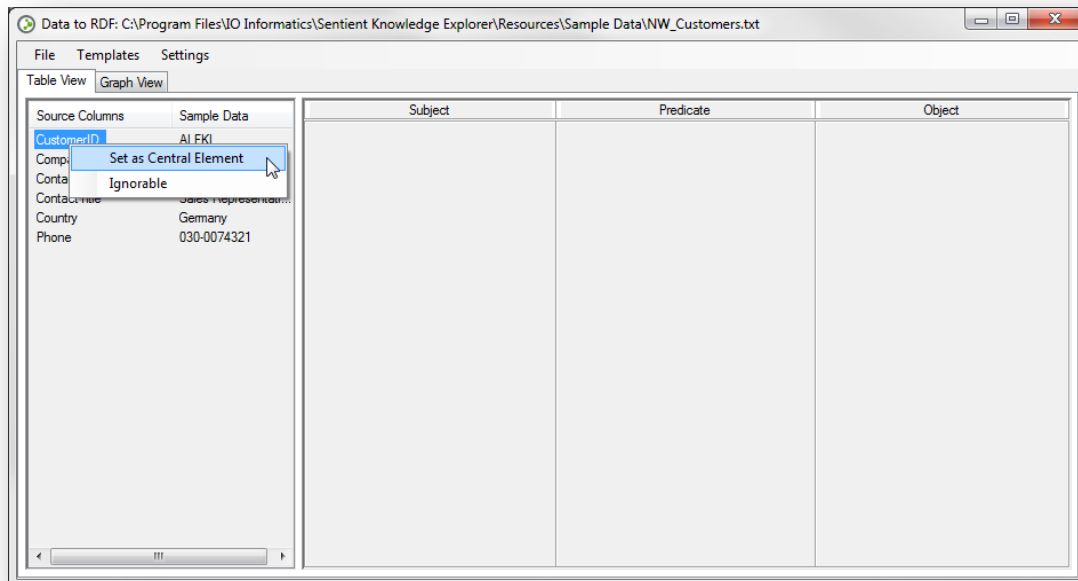
To load the file, launch the Knowledge Explorer and select “**File>Import>Spreadsheet or Delimited Text**”. Select the file and click on “**Open**”.



The Import Mapper Dialog Window

When creating a Knowledge Base with the Knowledge Explorer you are basically creating a series of statements, each consisting of a Subject, Predicate and Object, in the Resource Description Framework (RDF) as well as assigning a number of the elements to classes in an Ontology. (This is where the standards created by the Tim Berners-Lee and the W3C come in.) The Knowledge Explorer's import mapper dialog is used set up the translation of your data from spreadsheets, relational databases, etc. to RDF and thus create a Knowledge Base.

The import dialog displays column headers and a sample value for each column in your document. Right-click on any "Source Column" to set it as the Central Element for each row of the document. The Central Element is the equivalent to the primary key in a relational database table and serves as the Subject of the relationships you are creating from the spreadsheet. For this example, select "CustomerID" as the Central Element as pictured in the figure below.



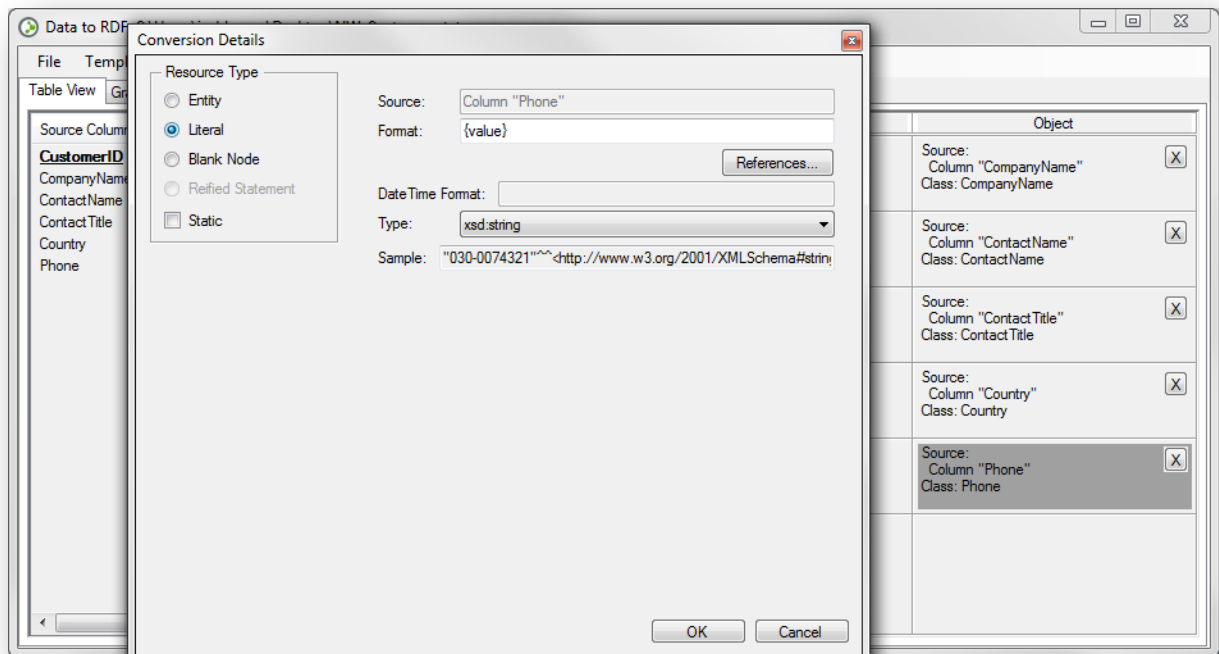
Set the Central Element for the data source

Once the Central Element has been set, select "Generate Templates" from the Templates menu. For each column in the spreadsheet, a relationship will be created using the Central Element as the subject and each of the other columns as the object of the relationship. These lines are called "templates". KE will use the column headers as class names as classes for each column. Predicates names will similarly be inferred from the class name of the objects.

For many applications, this is sufficient to begin to explore your data. However, the class names, entity namespace, and predicates can be further modified. To do so, double click on any cell to open up the configurations details menu.

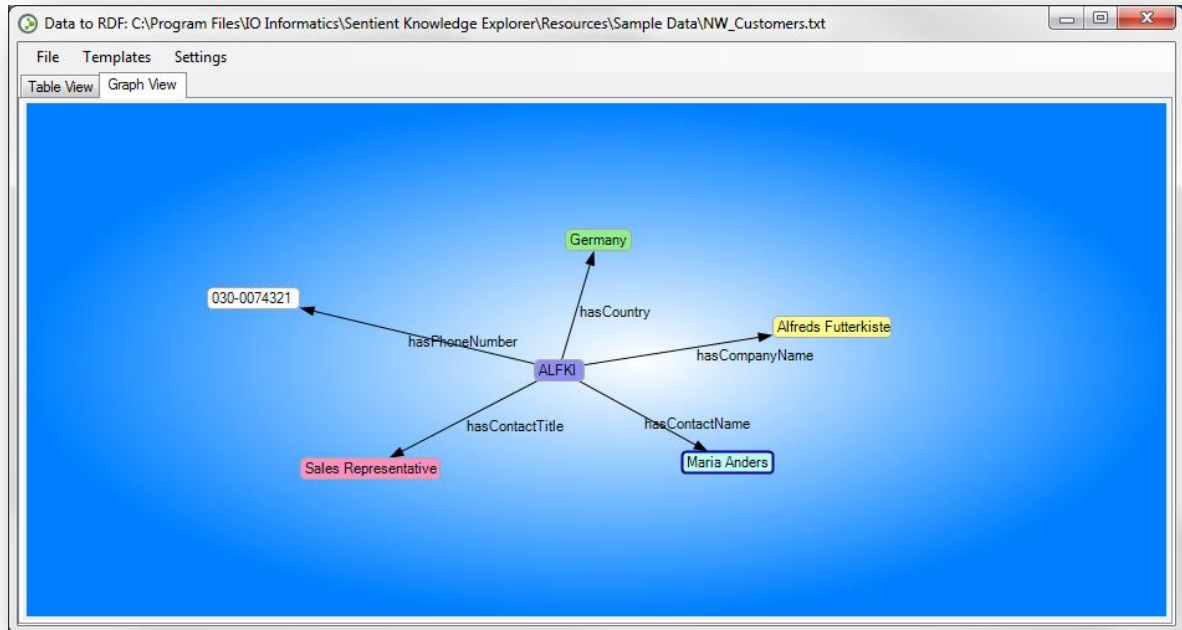
Try modifying the first predicate from “hasPhone” to “hasPhoneNumber”. Note that KE requires that predicates be a valid URI (e.g. <http://io-informatics.com/rdf/hasSomePredicate>) even though only the suffix (e.g. hasSomePredicate) will appear in the label. Changes made that do not result in a valid URI will not be saved.

Some data are better represented as literal values instead of entities with URIs. This is common when dealing with numerical values, dates and some text. These cases are sometimes referred to as properties of an entity. Double-click on the “Phone” class cell. In the “Conversions Details” window, select “Literal” from the “Resource Type” list on the left side of the window.



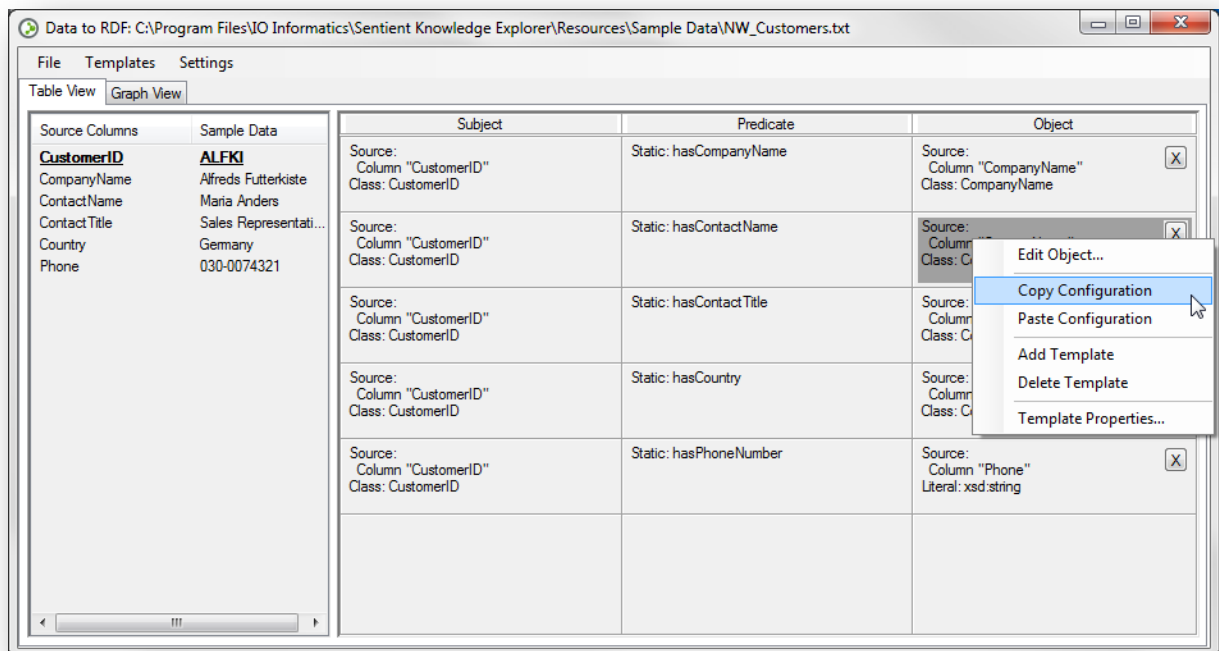
Change data from Entity to Literal data type

Click on the “**Graph View**” tab. The graph view displays the relationships between classes and literals as they will be imported into the graph. This will serve as a preview of the configurations you have made.



Graph View tab previews import configuration

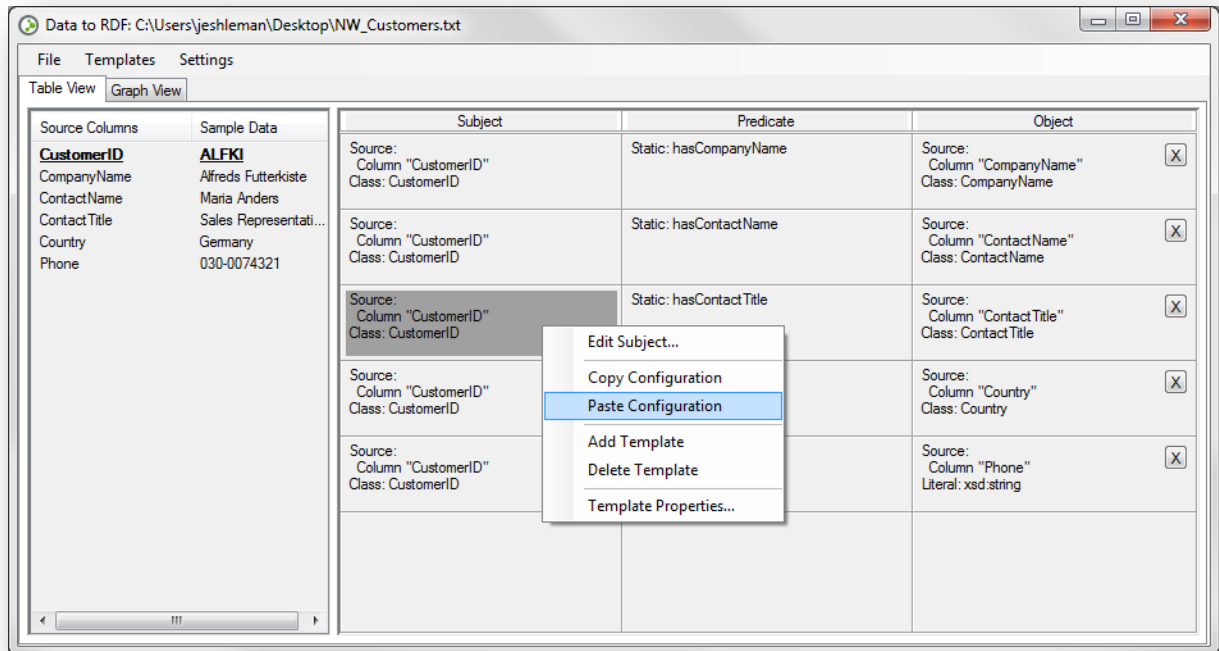
Return to the **“Table View”** tab and right-click on the cell designated **“ContactName”** in the **“Object Column”** and select **“Copy Configuration”** from the context menu.



Source Columns	Sample Data	Subject	Predicate	Object
CustomerID	ALFKI	Source: Column "CustomerID" Class: CustomerID	Static: hasCompanyName	Source: Column "CompanyName" Class: CompanyName
CompanyName	Alfreds Futterkiste	Source: Column "CustomerID" Class: CustomerID	Static: hasContactName	Source: Column "ContactName" Class: ContactName
ContactName	Maria Anders	Source: Column "CustomerID" Class: CustomerID	Static: hasContactTitle	Source: Column "ContactTitle" Class: ContactTitle
ContactTitle	Sales Representati...	Source: Column "CustomerID" Class: CustomerID	Static: hasCountry	Source: Column "Country" Class: Country
Country	Germany	Source: Column "CustomerID" Class: CustomerID	Static: hasPhoneNumber	Source: Column "Phone" Literal: xsd:string
Phone	030-0074321			

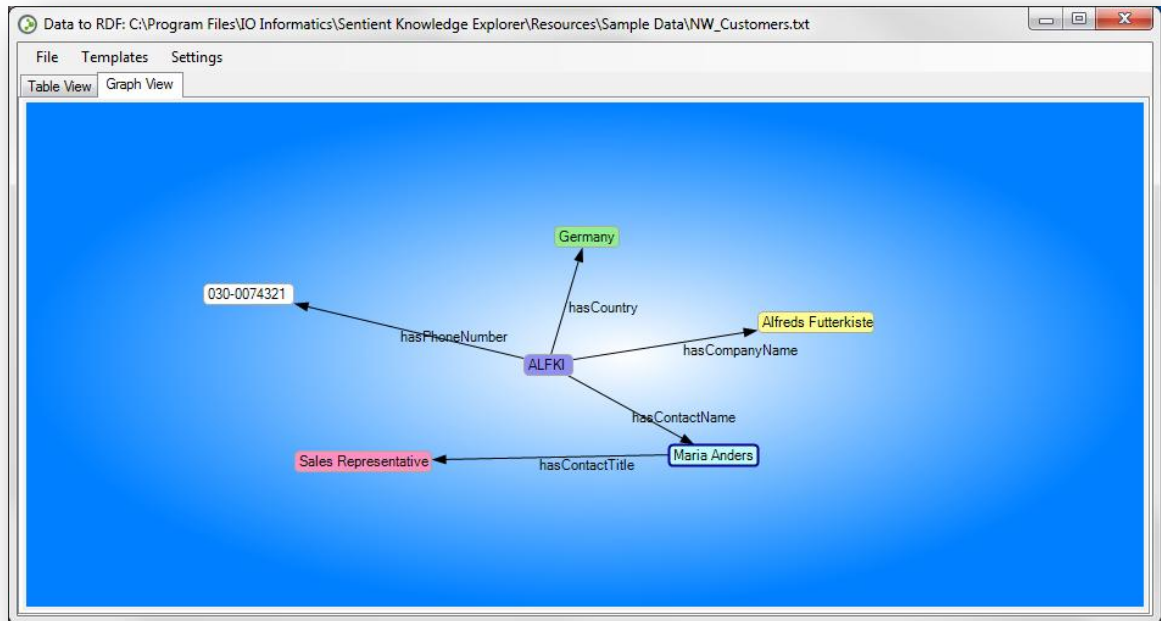
Copy Configuration

Right-click on the subject cell in the template line with the predicate “hasContactTitle”. Select “Paste Configuration” from the context menu as depicted in the image below.



Paste Configuration

After pasting the configuration to the new cell, select the “Graph View” tab. Notice how the entity “Sales Representative” now branches off of the “ContactName” rather than the CustomerID.




Graph View tab with new data relationship

Once you have completed your mapper, you can save it for future use. Mappers can be saved by selecting **“Save Mapper”** from the **“File”** menu. Previously saved mappers can be loaded with the **“Load Mapper”** option in the **“File”** menu. In this manner, one can easily load settings and class names for spreadsheets with the same column headers. Save the mapper and then click on **“Import”**.

Multiple files and datasets can be imported sequentially. Data saved in memory will be combined with subsequent imports until the Knowledge Explorer is closed or the memory is cleared with the **“File> Clear”** menu. With subsequent imports, new data can be added to existing classes, new classes can be created or some combination of these two can be used to build large linked data sets with the Knowledge Explorer.

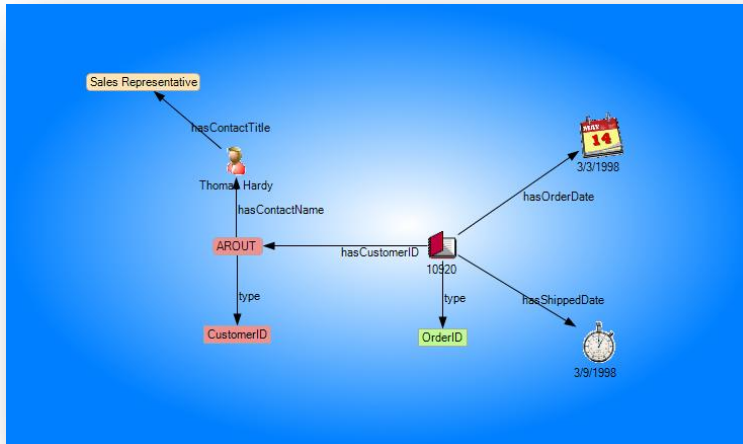
Select **“File> Import> Spreadsheet or Delimited Text”** and navigate to the file **“NW_Customers.txt”** once more.

Previously, we generated templates utilizing all columns of the spreadsheet. However, it is not necessary to use all columns. To add a single triple, select **“Add Template”** from the Templates menu. The source columns for any variable entity can be set by selecting the source column and dragging it into the appropriate cell of the Template (Triples) Configuration section. After adding a template, drag the Contact Name column into the **“Subject”** cell and the Phone column in as the object of the relationship. Modify the predicate to show the relationship as **“contactHasPhoneNumber”**.

Additional template lines can be added by selecting “Add Template” again. If you decide not to import a row, click on the  icon at the right of the particular line to remove it.

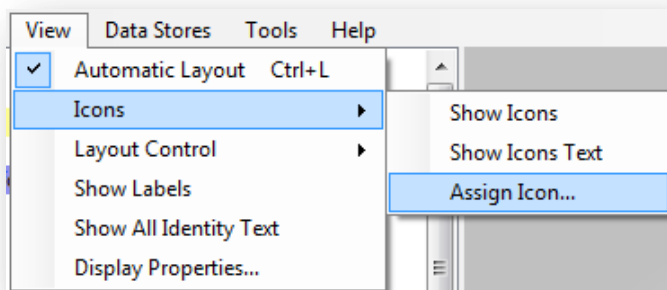
Customizing Your Display with Icons

The IO Informatics Knowledge Explorer includes the ability for users to associate icons with classes in their data set. Using icons to represent nodes in the graph simplifies the look of your knowledge base, allowing users to more readily recognize elements in the network and understand the big picture of what it represents.



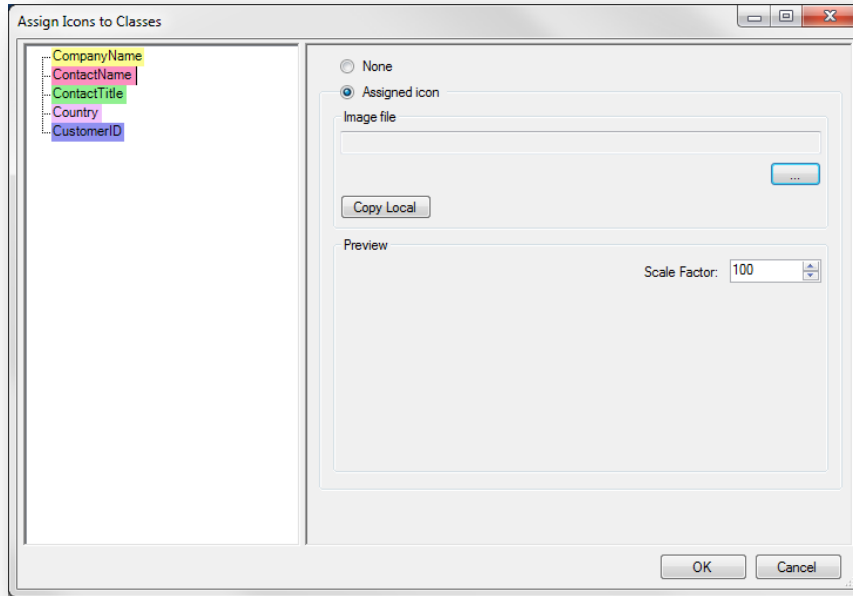
To associate icons with various classes in your data set, first import your data, if it is not already in memory, then make the associations as described below and save the file with the new icons.


Icon functionality can be accessed from the **View** menu bar:



From this menu you have the ability to toggle the icons on and off from the **Show Icons** selection. This provides the ability to see the full text term that the icon represents. You can also toggle on and off **Show Icon Text**, which provides a text sub-heading under the icon, referencing the actual entity.

Select “**View>Icons>Assign Icon...**” to choose, edit and refine the icons that will be used.



This menu lets you assign icons to any feature already loaded in the Knowledge Explorer. All you have to do is select a class on the left hand side, press the  button on the right, and select the specific image file you wish to use. A number of icons are installed with the Knowledge Explorer and are located in the directory:

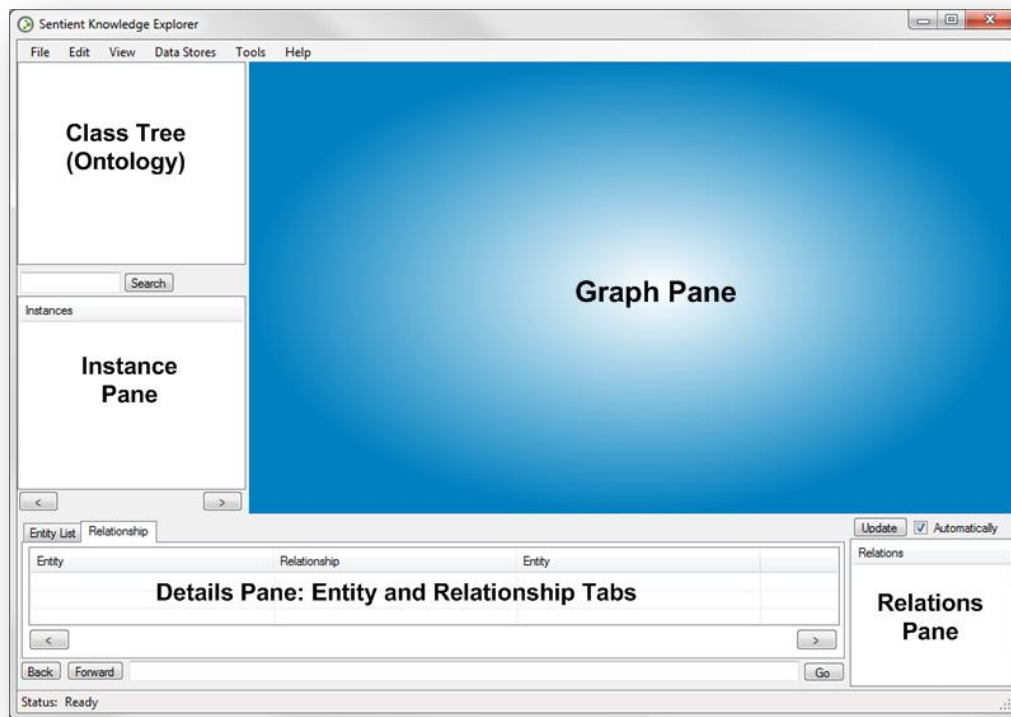
\Program Files\IO Informatics\Sentient Knowledge Explorer\Icons

In addition to these icons, the Knowledge Explorer supports many saved image formats (e.g. .jpg, .gif, .png, .bmp) and can use most image files as icons. For any icon you select, you can also adjust the Scale Factor of each icon using that feature below the preview. Once you have selected icons for each feature, press OK. Turn on the “Show Icons” option from the “View” Menu.

Saving Datasets

Once a dataset has been loaded from a spreadsheet or spreadsheets, it can be saved to a RDF file (in .n3, .nt or rdf/xml formats) to archive the records and make them available for easy access in the future. Data can also be augmented by imported an existing RDF file by simply dragging the file onto the Knowledge Explorer desktop.

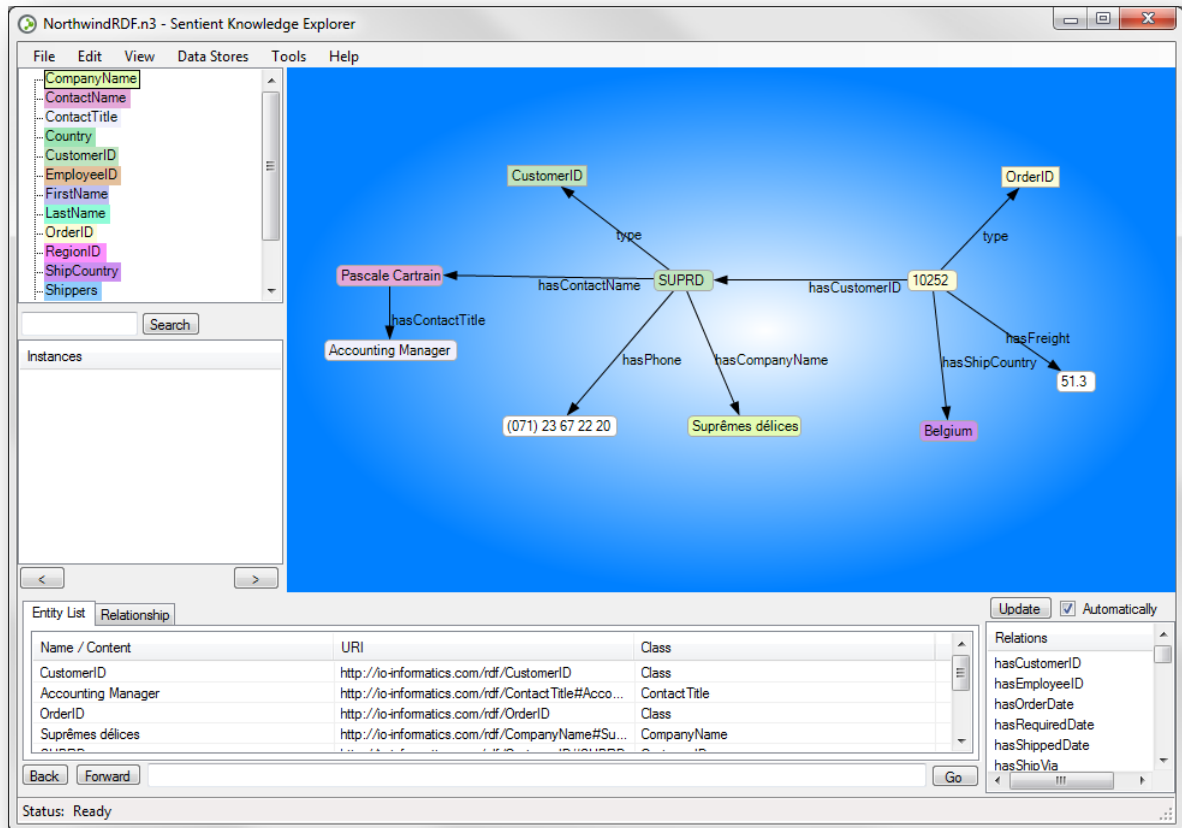
Now that you've created and saved your new data set, let's take a moment to review the Knowledge Explorer's interface. The classes you created during import will be visible in the Class Tree at the top left of the interface. By selecting any of the classes, all instances of that class will appear in the Instance Pane. Double clicking on any instance will bring that instance or node into the main Graph Pane, where one can explore connections between nodes graphically. To begin exploring your data, right click on any node in the browser and select "Show Connections>Single Entity..." and select any of the other data nodes connected to this node. Exploring the connections between nodes is one method of uncovering patterns within the data. The process of exploring data in the browser will be covered in more details in the following section: [Working with existing semantic data files.](#)



Data imported into the Knowledge Explorer, be it from spreadsheets, delimited text, existing RDF documents, or any combination of these sources can also be queried using the Knowledge Explorer's Query Tool. The Query Tool allows users to search a dataset based on the types of relationships between data as well as the characteristics of particular nodes. Querying data in this manner is a method of testing hypotheses about your dataset based on the pattern of connections between components of the dataset. A specific query example will be covered in the section [SPARQL queries](#) later in this document.

Working with existing semantic data files

Open the Knowledge Explorer and import the file NorthwindRDF.n3. It can be imported with the “File>Import>Semantic File...” menu, or by simply dragging the file onto the Knowledge Explorer window.

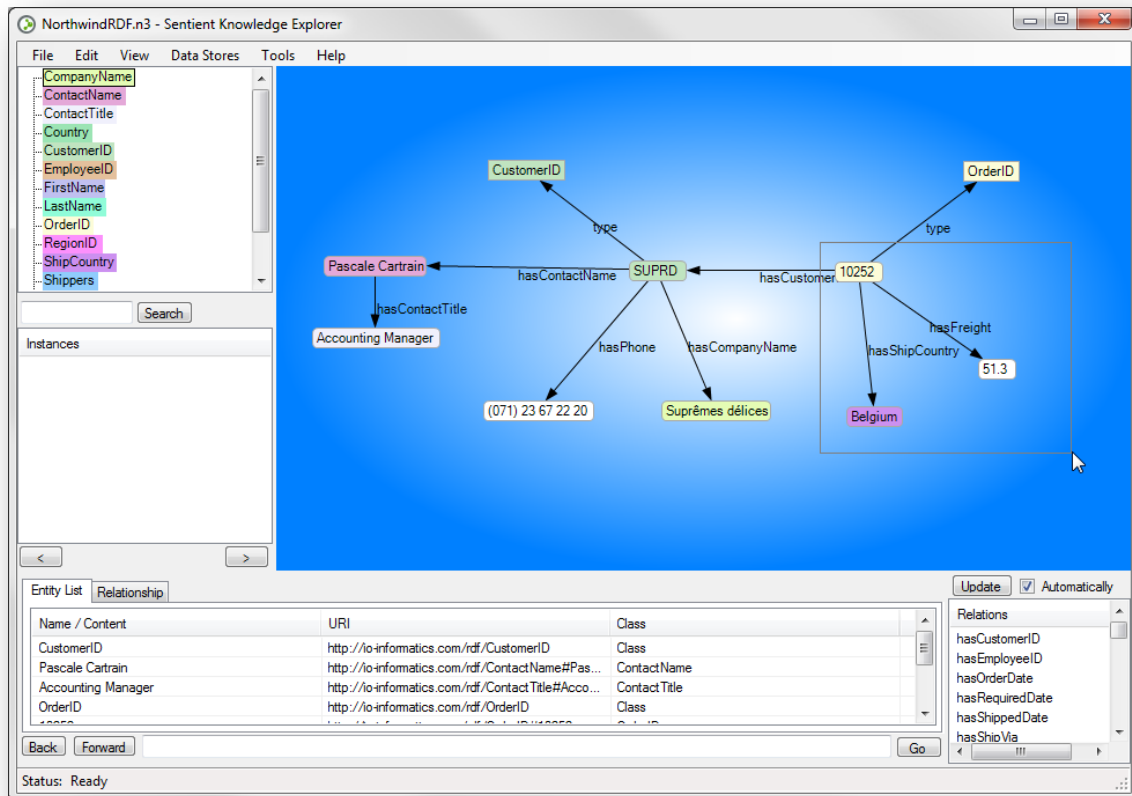


Clear the graph by right-clicking anywhere in the graph panel and selecting “**Clear Graph**” from the context menu. This will clear the graph, but all connections and data will remain in memory. Select “OrderID” in the Ontology browser pane (upper left) and select the first order, designated number **10248**. This node or entity will appear in your main layout display. Right-click on the entity and select “**Show connections > Select from list...**”. The Knowledge Explorer will display a list of all statements (entities and properties connected to the specific order ID) for the particular entity selected. Select a few of the statements and click on “OK”. The statements will be displayed in the main graph layout. This process allows us to use the Knowledge Explorer to rapidly view relationships between entities within your dataset and select commonalities and patterns within our data.

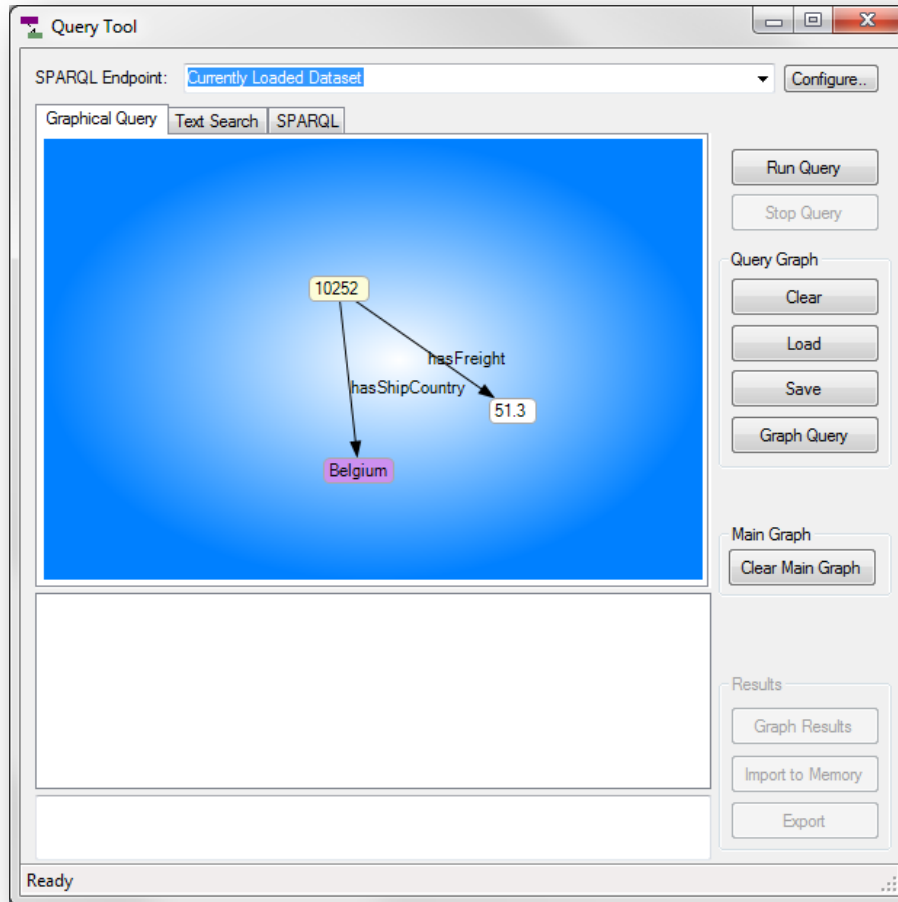
SPARQL queries

The Knowledge Explorer's query tools feature allows a user to construct complex, parameterized queries on semantic data sets using the SPARQL. In the previous exercise, we used the Knowledge Explorer view entities connected to a certain OrderID. The Knowledge Explorer's Query Tool can be used to construct more complex queries. In situations where the amount of data you are searching for is larger, it is more efficient to use the Query Tool to return results in tabular format. The Query Tool also lets us parameterize searches to customize result sets.

To demonstrate this process, start the Knowledge Explorer or, if it is already running, clear the memory with the **"File>Clear"** menu option. Load the file "NorthwindRDF.n3" once again. Select the three entities "Belgium", "10252" and "51.3" as depicted in the image below. With all three selected, right-click on any of the three and select "Create Query" from the context menu. A new window containing the query tool will appear with the selected items.

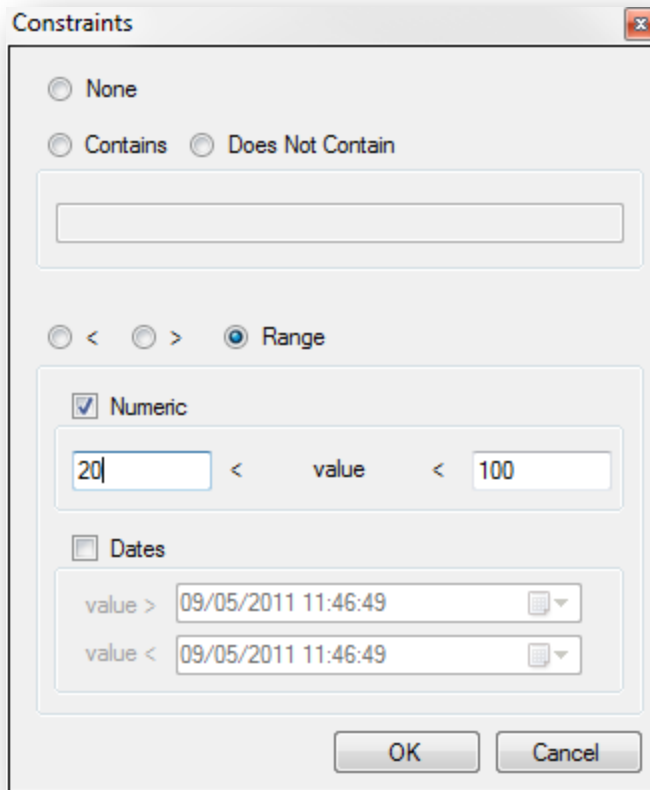


Select Entities in the graph layout



Query Tool window

Items displayed in the query tool menu will serve as the foundation to make a query. Right click on the OrderID (10252) and select “Make variable...” from the context menu. Right-click on the freight weight 51.3 and select “Restrict > Range”. Set the numeric range for values between 20 and 100 as depicted in the image below. Click on “OK” to close the “Constraints” window.



Set constraints for the shipping weight

Click on the “Run Query” button to return a result set containing all orders shipped to Belgium with a freight weight between 20 and 100kg.

All queries are converted into SPARQL query text. At any time, you can inspect the actual text by selecting the “SPARQL” tab. This allows you to copy and paste the text to run the query against other SPARQL end-points or to edit the query text prior to running it.

Next Steps

As mentioned in the introduction, you can keep adding data to your knowledge base to create ever richer data sets. Other techniques not covered in this tutorial are the use of Thesauri, storing the data in a centralized database for sharing with your colleagues, using automated reasoning, linking to external data sources, and more. You can find additional information in the User Manual, the technical support forums at forums.io-informatics.com. IO Informatics also provides training and professional services around both our products and the Semantic Web in general to help you meet your goals.