
Article

[Benjamin De Boe](#) · Jun 7, 2016 6m read

[Open Exchange](#)

iKnow demo apps (part 2) - Set Analysis demo

This is the second article in a series on iKnow demo applications, showcasing how the concepts and context provided through iKnow's [unique bottom-up approach](#) can be used to implement relevant use cases and help users be more productive in their daily tasks. Last week's article discussed the [Knowledge Portal](#), a straightforward tool to browse iKnow indexing results.

This week, we'll look into the Set Analysis demo, a slightly more advanced application where you'll be using the concepts identified by iKnow to organize your content into sets of documents. The original version of this demo was developed by Danny Wijnschenk & Alain Houf for an academy session at GS2015, but the app has evolved significantly since then.

Installation and setup

This demo is available for download from a [GitHub repository](#), where you can either check it out using your preferred Git client, or download it as a zip outright. You should be able to import and compile all classes in any release of Caché starting with 2015.2.

The application consists of a single (CSP) page with AngularJS-infused HTML and associated JavaScript code that calls a REST interface. To make the application work, you'll need to set up a web application that redirects to the Demo.SetAnalysis.RestHandler class, either manually or through the Demo.SetAnalysis.Utils.CreateRestWebApp() method included in the package.

The demo is meant to work with any domain without prior setup, such as the Aviation Events demo domain discussed in the previous article, or one you define all by yourself [as described here](#). In this article, we'll use another demo domain for which the definition is included in the GitHub repository in the Demo.Hotels.DomainDef class. For copyright reasons, we unfortunately cannot publish the data, but you can populate the Demo.Hotels.Review table manually with a few reviews of your own or use a tool like [import.io](#) to scrape some from the web. If you have a table with a text column, but haven't yet created an iKnow domain for it, you may want to take a look at the Demo.SetAnalysis.Utils.Setup() method's class reference to automate that task.

Exploring the domain

To access the demo, first note the domain ID of the domain you wish to work with and then access the following URL: <http://localhost:57772/csp/setanalysis/SetAnalysisDemo.csp?1>. In my case, I have a namespace and web app named "setanalysis" and the domain ID is 1. Depending on the data you loaded, you should see something like this:

Table Concepts Sets Sentiment						
convenient						
Author	SourceId	StayMonth	StayType	FullText	Sets	
Mainiac817	368	November 2007	family	Nice My daughter and I stayed here for 2 nights while touring Boston. it was very nice . The beds were very comfy and the staff very helpful . Parking was expensive but i figured as much.	parking	
blessedmomofour	361	September 2007	family	Very Satisfied We stayed at the Marriott Cambridge while moving my son into Harvard, and found it very convenient. The parking was a problem, since the garage was filled, but we were able to park right in front. The room was very clean , and the bedding was comfortable .	parking	
e j	360	May 2015	friends	... we were in Cambridge for a MIT and had not been pleased last year at the Kendall Hotel Rooms are a little dark as the windows are just in one corner. but we were on the 17th floor and had a decent territorial view, although we didnt pay extra for the view of Boston skyline, which did look amazing from the elevator lobbies . The guest closet was annoyingly small and nearly blocked by ridiculous bi-folding doors. ... Two stop to Harvard Square and lots of restaurants, although there are nice choices within walking distance	parking	
Aimee B	359	May 2015	solo	... I was in Boston for a couple of days (a Saturday and a Sunday) to visit friends before heading up to Maine and chose this hotel based on location and price. ... The hotel is not in the middle of Boston , but a very quick and easy T ride away. ... Not very professional and not very customer friendly in my opinion, especially for the price per day Overall, I paid a decent price for the hotel and was satisfied with my stay. ... I wanted to stay at this hotel again but when making reservations the previous month, I found that the prices for the Thursday following my stay doubled. ...	parking	
Myrolexs	356		unknown	Good location, quiet at night This hotel is located in a good location in Cambridge , just a short walk or ride on the Red Line to Beacon Hill The room had the amenities one would expect from a Marriott, not too much but all that is needed. The bar/restaurant was very good with a varied food and drink menu . The staff overall from the front desk to the parking attendants were very	parking	

The first tab of the app provides a simple tabular view of the hotel reviews in our domain, with a summary of the text displayed and negation underlined and sentiment (see below) highlighted in green and red. You can click on a row to display the entire record. While helpful to get a first glimpse of what's in your records, the goal of iKnow is not having to read through hundreds of documents, so let's move on to the "Concepts" tab.

Table Concepts Sets Sentiment						
show all						
Concepts						
hotel	466					
room	301					
boston	180					
mit	118					
rooms	113					
location	98					
cambridge	90					
staff	86					
lobby	83					
marriott	82					
clean	75					
right	72					
great location	63					
starbucks	62					
great	55					
Similar Entities CRCs						
room	301					
rooms	113					
room service	15					
clean rooms	8					
nice room	5					
every room	4					
clean room	3					
corner room	3					
each room	3					
hotel room	3					
two rooms	3					
breakfast room service	2					
comfortable room	2					
comfortable rooms	2					
Sentences						
This is a high rise downtown hotel, clean room , small bathroom , great views and a great location .						
Good sized, clean room and this time with an excellent view of the Boston skyline.						
Pros: great location right next to the T station and MIT, clean room with a living area (sofa and a coffee table), cool corner windows, toilets that flush very quietly (some hotels have violently flushing toilets), very friendly lobby staff , comfortable beds and covers , easy self-checkout using TV (they slip the receipt under your door in the morning).						

This interface will look somewhat familiar to the Knowledge Portal we discussed last week. Let's go through the widgets one by one. The Concepts widget displays the top concepts for your domain, sorted by frequency. When you select a concept from the list or enter a value in the "concept fragment" right above the central widget, the Similar Entities list will be populated with all entities similar to that fragment. You can also open a tab displaying CRCs containing the seed entity in the same widget. Clicking on an entity or CRC in this list will populate the third

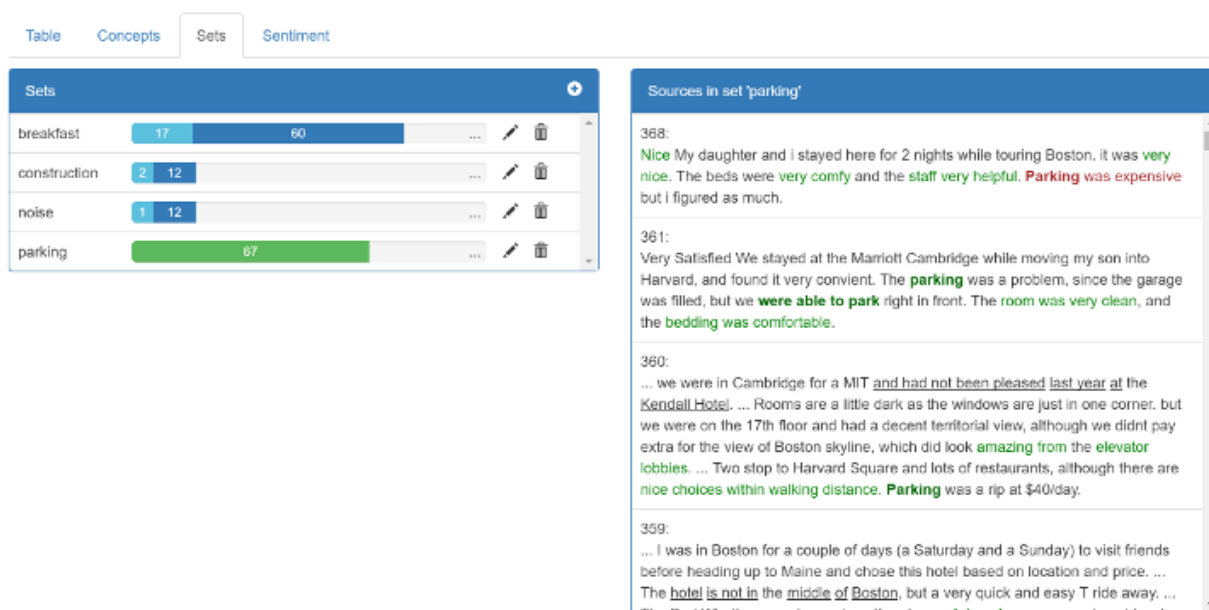
widget with all Sentences containing that element, to give you a sense of context. In the screenshot, you can see how iKnow's entity level already allows you to separate occurrences of the word "room" into cases where the author meant an actual room or a different concept like room service.

Creating simple sets

Now, to start organizing our data, we can select a number of terms (CRCs and/or entities) that describe a particular theme or topic and save those as defining a set, using the input field and button above the sentences widget. Successive "save" commands will append to a set with the given name if it exists, so for example you can create a set combining relevant terms similar to "noise", "loud" and "noisy".

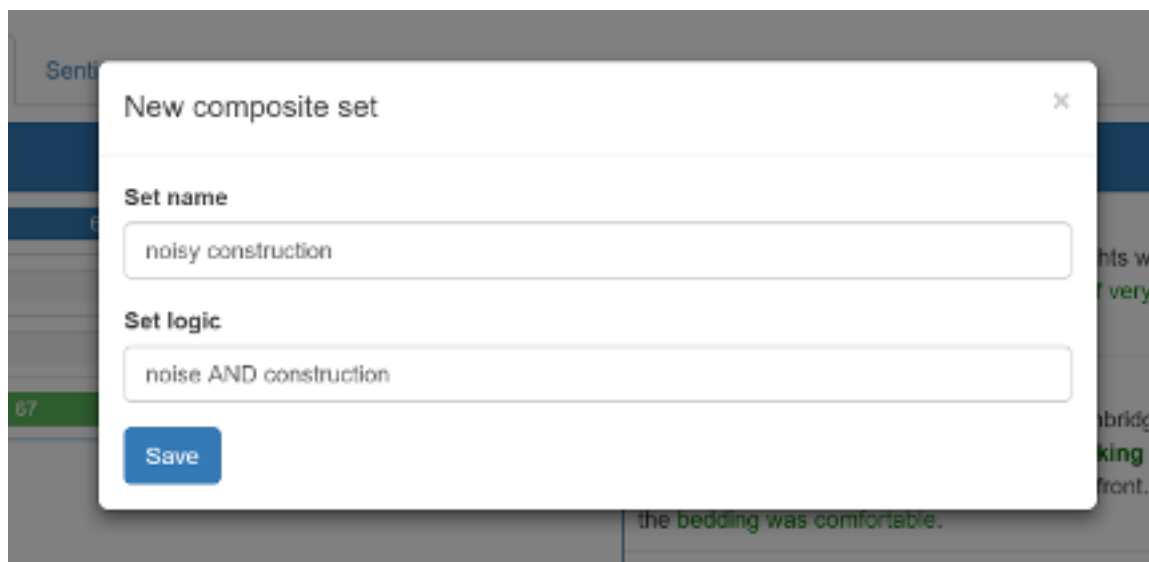
Once you've defined any sets, you can use the dropdown right above the concepts widget to limit all results to a particular set. The sunglasses button right next to it allows you to toggle any blacklists defined in your domain on or off, to filter specific terms from the results. This button will be hidden if your domain does not contain any blacklists.

Let's turn to the next tab to zoom in on our sets.



The left widget displays an overview of the sets you've created thus far and how many records they consist of, visualized as a bar. As soon as you select a set, each bar will be split into the records that are shared with the selected set, and the ones that are only in the bar's set (cf tooltips). On the right side, you'll see the actual records in the selected set, showing a summary of the text at least including the sentences with the key terms defining the set. If you click a bar fragment that represents the overlapping records, the widget on the right will only display those records in the intersection.

Often you'll be interested in more complex combinations of sets, rather than simple overlap. To create a composite set, you can click the plus icon in the left widget and specify the sets that should be combined through boolean AND/OR logic.

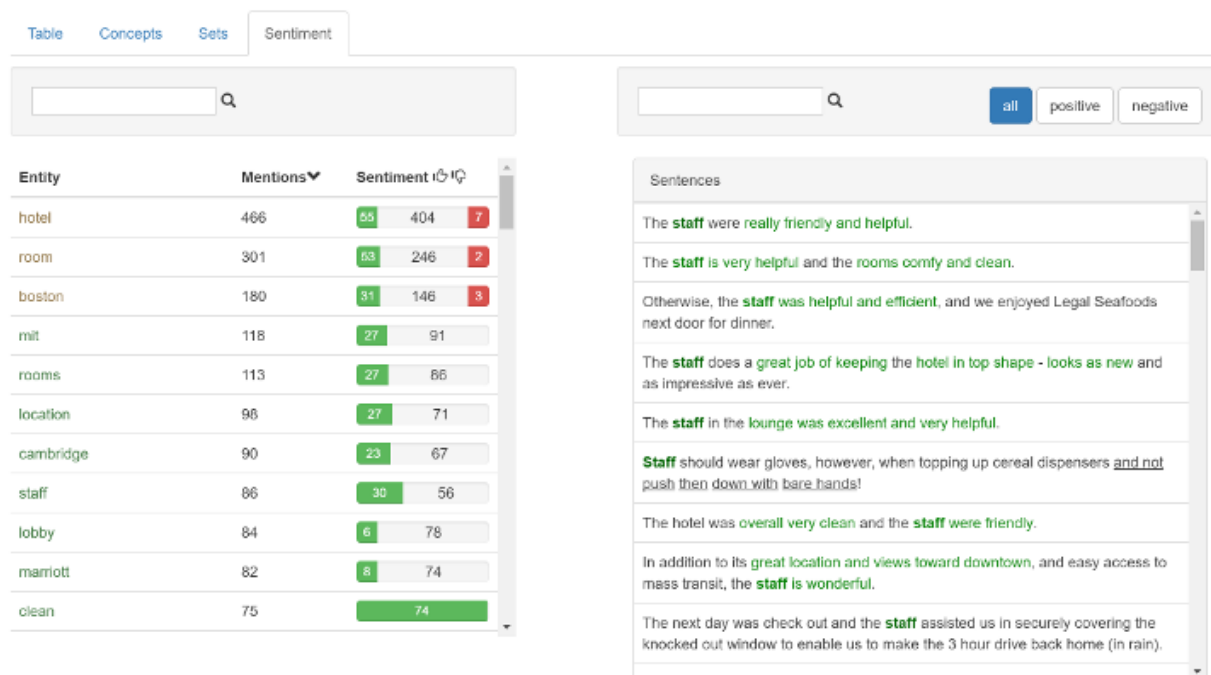


If you click the pencil icon next to the bar, you'll be redirected to the entities tab, where you can edit the list of terms defining the set, removing the obsolete ones. Doing this for a composite set will take you back to the dialog, where you can change the logic.

While the data used here makes for a bit of a toy example, a more serious use case for which we've been using this demonstration is for [patient cohort selection](#). In clinical trials for a new medication or treatment, billions of dollars are spent on conducting tests with an appropriate group of candidate patients. However, finding the right group of candidates is hard, as the criteria for inclusion are typically complex and highly selective. More often than not, these criteria include constraints on elements that may not be available in a nicely structured column you can include in a SQL query's WHERE clause, but buried deep down in the clinical notes of a patient's file. Therefore, tools that help you in sifting through these vast volumes of unstructured data and allow you to base patient cohort criteria on both structured and unstructured data, as displayed here, are highly valuable and can save hospitals and clinical research organisations millions of dollars.

Browsing sentiment data [optional]

For a particular customer demo, this app was extended with a fourth tab allowing you to browse sentiment data. This tab is only visible if your domain contains actual sentiment data (as does our Hotel review domain), which requires a user dictionary specifying sentiment terms as explained in [a separate article](#). It's not particularly integrated with the rest of the Set Analysis app, but it might fit into your story so we're leaving it for your convenience.



This article describes a first "external" demo application we've been using extensively on the road. Next week, we'll look into a second one for building dictionaries based on an iKnow domain, in sort of a bottom-up fashion. While we try to keep them generic and reusable, all of the demos in this series gradually implement more complex scenarios that get closer to the use cases we're seeing at customers. While it's obviously our goal to work with partners to implement these full-fledged solutions, often the steps towards them and limited applications built along the way offer helpful capabilities too.

Feel free to comment, suggest further extensions (or implement them yourself!) or point us to any inaccuracies or bugs!

[#InterSystems Natural Language Processing \(NLP, iKnow\)](#)
[Check the related application on InterSystems Open Exchange](#)

Source URL: <https://community.intersystems.com/post/iknow-demo-apps-part-2-set-analysis-demo>