Innovations by **InterSystems**



iKnow and DeepSee

Agenda

- What is iKnow?
- Semantic Analysis.
- %iKnow.Queries
- Matching Dictionaries.
- %iKnow.Semantics.

Newer features:

- Attribute Customizations.
- iFind.
- Text Categorization.
- iKnow features in DeepSee.
- Configuring iKnow and DeepSee.

What is iKnow?

iKnow is a semantic analysis tool.

- Indexes the concepts and relations within text for querying and analysis.
- Uses language models rather than training data or ontologies to detect relations.
- Supported languages: Dutch, English, French, German, Portuguese, Russian, Spanish, Ukrainian, Swedish*, and Japanese*.
- Sources of text include: Plain text files, SQL fields, social media.

*Support added in 2016.1 release.

Semantic Analysis: Relations, Concepts, Negation

 The patient suffered from acute hypertension but did not mention any chest pain.



iKnow and DeepSee Slide 4

Semantic Analysis Results

Concept

patient

acute hypertension

chest pain

Relation

suffered from

but did not mention

Concept-Relation-Concept

patient suffered from acute hypertension

acute hypertension *but did not mention* chest pain

iKnow and DeepSee Slide 5

Importance of Language Models

- iKnow indexing is subject matter neutral.
 - A language model applies to any text written in the language: medical, legal, scientific, business, and so on.
- iKnow indexing automatically detects meaningful word groups.
 - Labels "acute hypertension" and "chest pain" as concepts.
 - Labels "but did not mention chest pain" as a negation context.
- No need for ontologies or training data.

%iKnow.Queries

Includes:

- GetTop() Most frequently occurring entities across a set of sources.
- GetRelated() Entities in a relationship with the supplied entity.
- GetByEntities() All CRCs or paths containing a particular set of entities.
- GetSummary() Most relevant sentences in a source.
- GetSimilar() Entities similar to a given entity.

Innovations by **InterSystems**

iKnow and DeepSee Slide 7

Matching Dictionary

User provided group of related terms.

- Provides external (domain) knowledge to iKnow results.
- Allows for coarser grained analysis.
- Example (2001 A Space Odyssey): hal □ hal. hal9000 □ hal. heuristic algorithm computer □ hal.
- iKnow smart matching mechanism returns a match score.
 - Configurable threshold for matches.

iKnow and DeepSee Slide 8

iKnow Architect (2016.1)

- Management Portal Tool for creating, configuring, and managing iKnow domains.
 - Domain Settings, Metadata, Data Locations, Blacklists
 - Compile and build domains.
 - Launch indexing and knowledge portal pages.
- Some iKnow features not supported by Architect. Edit class definition using IDE.
 - Matching Dictionaries.

Demonstration

.kn w								
:FILE:c:\deepsee\know\movies\star wars episode v - the empire strikes back.txt	movies •							
sentence: 1 to 59 display!								
This is a sample User Interface built on top of the iKnow APIs, visualizing the results of the iKnow Smart Indexing API. You can either select an existing source from the dropdown list, or use the input button to enter free text directly.								
bide								
Indexed sentences								
The opening crawl reveals that the Galactic Empire has pursued the Rebel Alliance across the galaxy, forcing them to establish a secret base on the remote ice planet Hoth.								
while Luke is patrolling paar the base he's knocked unconscious by an indigenous predator, the Wampa								
Back at the base, the smuggler-pilot Han Solo (Harrison Ford) announces his intention to leave the Rebels and pay the debt he owes to the gangster Jabba the Hutt, much to the displeasure of Princess Leia (Carrie Fisher).								
After Han discovers that Luke has not returned from patrol, he delays his departure and leaves the base to search for him.								
Luke escapes the Wampa's lair but is overcome by the cold.								
He sees an apparition of his late mentor, Jedi Master Obi-Wan Kenobi (Alec Guiness), who instructs him to receive training from Jedi Master Yoda (voice: Frank Oz) on the planet Dagobah.								
Han finds Luke and gives him shelter until they're rescued the following morning.								
When ground scans detect an object outside the base perimeter, Han and Chewy investigate, and find an Imperial probe droid that transmits the location of the Rebel base to the Imperial fleet before firing upon Chewy and being destroyed.								
On the command ship of Darth Vader the droid's transmission is picked up by Captain Piett, but his superior, Admiral Ozzel, dismisses the message until Vader sees the transmission and realizes it is the Rebel base.								
He overrules Ozzel and orders the Fleet to the Hoth system.								
Later, as the Fleet emerges from trans-star warp, Vader is informed by his infantry commander, General Viers, that the Rebels have set up infantry trenches and an energy shield to protect them from the Empire's orbital bombardment.								
Ordering Viers to launch a surface attack, Vader is furious that Ozzel has clumsily given away surprise, and he communicates this displeasure to Captain Piett as he promotes him on the spot to Admiral while Ozzel is given Vader's own fatal brand of reprimand.								
The Imperial forces land their ground assault walkers beyond the energy shield and Luke leads his squadron of flying speeders into battle.								

Innovations by **InterSystems**

iKnow and DeepSee Slide 10

%iKnow.Semantics (2012.2+)

Introduces concept of *dominant* entities.

- Most important entities not most common.
- Algorithm revised for 2015.2 release.
 - Explained in documentation.

Includes queries:

GetBySource() – Dominant elements in a specific source.

 BuildOverlap() – Generates dominant term overlap information for all sources in a domain.

- FindMostTypicalSources() Most typical sources.
- FindBreakingSources() Most atypical sources.

iKnow and DeepSee Slide 11

Attribute Customizations

Negation.

 Augment default markers with additional markers for particular use cases.

Sentiment.

- No default markers.
- Supply custom sentiment markers.
- Attribute markers.
 - Supply custom markers in User Dictionary.
 - iKnow performs attribute tagging during loading.

iFind

SQL feature for performing text search.

- Add iFind index to columns containing text.
- Include iFind index syntax in WHERE clauses of SQL queries.
- Support for the following searches:
 - Stemming and de-compounding.
 - Word and word phrase search.
 - iKnow entity search.
 - iKnow semantic search using path, proximity, and dominance information.

Text Categorization

- Label (categorize) source texts based on their contents (entities and relations).
 - Create a classifier by analyzing an existing (training) set of already labelled texts
 - Apply classifier to new and as yet unlabelled texts.
- Wizards available for building and testing classifiers.
 - System Explorer
 iKnow
 Text
 Categorization

DeepSee and iKnow

 DeepSee cubes can include iKnow indexing results and analyses:

- iKnow Dimensions.
 - Entities (concepts and relations).
 - Dictionary matching results.
 - Use as rows, columns, and filters on pivot tables just like data and time dimensions.
- Detail Listings.
 - iKnow summaries.
 - Content Analysis Plugin to allow users to perform a variety of iKnow analyses on text sources.

Demonstration

Aviation event reports	× - + US Accidents by state			+ US Accidents by state x - + "Highest injuries" vs matching results								
Aviation event reports												
Filters	Event count Alabama Alabama Alabama Alabama Alabama Alabama Alabama Alabama California Colorado Connectic Delaware Florida Georgia Hawaii Idaho Illinois Indiana	Unknown ^ Alabama Alaska Arizona California Colorado =	Dictionary-			None		Minor	Serious			
Year Q Aircraft type			iKnov	Know matching results		e	342	30				
					min	or	2	115				
					seri	ous		1				
				fata	1							
		Connecticut	All reports			527	163	1				
		Delaware										
		Georgia	۲ III									
		Hawaii										
		U Idaho	X = + Top concepts in Event Reports									
		Indiana										
		lowa		antitu	fraguanau	aproad						
		Kansas	201	1 fotol injurios	requency	spread						
		Louisiana	20	vertical	0/	02						
		Maine	202	stabilizer	90	62						
		Massachuset	203	3 100 feet	70	61						
		Michigan	204	4 alaska	137	61						
		Minnesota	205	5 cruise flight	71	61						
		<										
			_									

Innovations by **InterSystems**

iKnow and DeepSee Slide 16

iKnow Dimensions

- Entity dimension.
 - Single level.
 - Members are entities (concepts or relations).
 - Analyzer displays first 100 in decreasing order by spread.
 - Filter options contain all entities. Searchable.
- Dictionary dimension.
 - Level 1: one member for each dictionary.
 - Level 2: one member for each item containing all matches for that item.
 - Matching dictionaries loaded as termlists.

iKnow Measure

Connects unstructured data to cube.

- Purely configuration. Not visible to Analyzer.
- Connects DeepSee cube to text sources and dictionaries.
- Referenced by iKnow dimensions.

Content Analysis Plugin

- Launch from Analyzer or Dashboard.
 - Select cell and click M
- iKnow features include:
 - Content Analysis.
 - Typical and breaking sources.
 - Entity Analysis.
 - Overview: frequency and spread for 10 most common groups.
 - Cell breakdown: distribution of entities selected on Overview tab.
 - Entities: frequency and spread for entities similar to entity selected on Cell breakdown.

iKnow and DeepSee Slide 19

Demonstration

Menu Home About He	lp Logout DeepSee > Architect Server,D User, U	ENELSONE6410 Namespa nknownUser Licensed	ace: SAMPLES Switch to: ISC Learning Services - Instructors Instance: CAC	CHE	DeepSee by InterSystems
View:	New Open Save Compile	Build Document	tation		Architect
Source Class	Model Elements Add Element Undo	o Expand All Collap	se All Reorder		»
▼ Aviation.Event	AviationEvents	Element Type	Details		Details Tools
☐ %ID	▼ Measures				Select an item to view details
Aircraft 2	SeatsTotal	number measure	Aircraft1.SeatsTotal	×	
AirportDirection	InjuriesTotal	number measure	InjuriesTotal	×	
AirportDistance	Report	iKnow measure	Narrative1.Report	×	F.
AirportElevation	▼ Dimensions				
AirportId	▼ AirportDim	data dimension		×	
	H1	hierarchy		×	
Altimeter	AirportName	level 1	AirportName	×	
City	▼ AirportLocationDim	data dimension		×	
CoordsLatitude	H1	hierarchy		×	
CoordsLongitude	AirportLocation	level 1	AirportLocation	×	
CoordsSource	▼ LocationDim	data dimension		×	
CountryCode	H1	hierarchy		×	
DayOfWeek	Country	level 1	Country.Name	× 0 0	
EventDate	State	level 2	State.Name	× o e	
L] EventId	City	level 3	City	× 0 0	
	▼ EventDateDim	time dimension	EventDate	×	
InjuriesGroundEatal	H1	hierarchy		×	
InjuriesGroundMinor	EventDateY	level 1	Year	× o e	
InjuriesGroundSerious	EventDateM	level 2 MonthNumber 🗶 0 0			
InjuriesHighest	▼ InjuriesHighestDim	data dimension		×	
☐ InjuriesTotal ☐ InjuriesTotalFatal ☐ InjuriesTotalMinor	H1	hierarchy		×	
	InjuriesHighest	level 1	InjuriesHighest	×	
InjuriesTotalNone	▼ LightConditionsDim	data dimension		×	
< •	H1	hierarchy		×	v.

iKnow and DeepSee Slide 20

Configuring iKnow Measure

iKnow Measure:

- Source Values: Property or expression.
- Aggregate: Count.
- Type: iKnow.
- iKnow Source: string, stream, file, or domain.
- Dictionaries: select from available termlists.

Configuring iknow Dimensions

Entity Dimension.

- Dimension Type: iKnow.
- iKnow Type: entity.
- iKnow Measure: iKnow measure name.
- Dictionary Dimension
 - Dimension Type: iKnow.
 - iKnow Type: Dictionary.
 - iKnow measure: iKnow measure name.

iKnow Listing Features

Include iKnow summary.

- \$\$\$IKSUMMARY[iKnowMeasure, summaryLength].
- Include content analysis plugin.
 - \$\$IKLINK[iKnowMeasure].
 - Allows users to see: summaries, dictionary matches, negation contexts, and dominant entities for selected source(s).

Suggested Reading

- Using iKnow.
- Advanced DeepSee Modeling Guide
 Using Unstructured Data in Cubes.

iKnow and DeepSee Slide 24



What are the key points for this module?



iKnow and DeepSee Slide 25