

Course: Information Retrieval

Lecturer: Alessandro Bozzon

Date: April 7, 2015

Team 7:

SKiIR

Explanation Retrieval

“Sophisticated Keyword Intelligent Information Retrieval”

Ruben Verboon

Herman Banken

Wiebe van Geest

Kors van Loon

Maarten Steinfort



Google's Nest to Buy Security Startup Dropcam for \$555 Million

by Brian Womack

6:01 AM CEST June 21, 2014



June 21 (Bloomberg) -- Google Inc.'s Nest Labs is acquiring Dropcam Inc. for \$555 million to boost its offerings for the connected home.

<http://www.bloomberg.com/news/articles/2014-06-21/google-s-nest-buying-security-company-dropcam-for-555-million>

Imagine...

You are reading an online newspaper about a new startup company that has been bought up by Google.



the world, an offering that would broaden Nest's product line-up into home security. Nest sells digital thermostats and smoke alarms that can also be checked and adjusted remotely from mobile gadgets.

"We care very deeply about helping people stay connected to their home, especially when they're not in their home," said Matt Rogers, co-founder of Nest, which is based in Palo Alto, California, in an interview. "This is another offering that helps people get insights into their home."

The purchase underlines how Nest, which Google bought for about \$3.2 billion earlier this year, is working to become a bigger player of connected devices for consumers. The trend is part of a technology movement dubbed the "Internet of things," where more gadgets and everyday items are connected to the Web and can deliver data and be controlled by mobile devices.

The products gather data on consumer habits and usage. Privacy experts have questioned how Google, the world's largest search engine, would use Nest's data. Nest has said its information will remain separate from its parent company.

The purchase underlines how Nest, which Google bought for about \$3.2 billion earlier this year, is working to become a bigger player of connected devices for consumers. The trend is part of a technology movement dubbed the “Internet of things,” where more gadgets and everyday items are connected to the Web and can deliver data and be controlled by mobile devices.

Wow, did I miss something?!

What is **Nest**?

Why did Google buy it for **\$3.2 billion** dollars?

And now they are buying another company for **\$555 million** dollars?

And.. what is the **Internet of Things**?

.. where more gadgets and everyday items are connected to the Web and can deliver data and be controlled by mobile devices.

What would you do?

What is the **Internet of Things**?



Quit the article?!

What would you do?

What is the **Internet of Things**?



Go to a competitor
newspaper website?!

What would you do?

What is the **Internet of Things**?

*Wouldn't it be nice if
someone could explain this
to you?*



Create account Log in

Article Talk Read Edit View history Search

WIKIPEDIA The Free Encyclopedia

Main page Contents Featured content Current events Random article Donate to Wikipedia Wikipedia store

Interaction Help About Wikipedia Community portal Recent changes Contact page

Tools What links here Related changes Upload file Special pages Permanent link Page information Wikidata item Cite this page

Print/export Create a book

Internet of Things

Internet of Things is that all user objects are embedded with software and are connected and interacting with each other.

The term "Internet of Things" was first documented by a British visionary, Kevin Ashton, in 1999.^[1] Typically, IoT is expected to offer advanced connectivity of devices, systems, and services that goes beyond machine-to-machine communications (M2M) and covers a variety of protocols, domains, and applications.^[2] The interconnection of these embedded devices (including smart objects), is expected to usher in automation in nearly all fields, while also enabling advanced applications like a Smart Grid.^[3]

Things, in the IoT, can refer to a wide variety of devices such as heart monitoring implants, implantable medical devices, home electrical appliances, mobile phones, laptops, automobiles with built-in sensors, or field operation devices that assist fire-fighters in search and rescue.^[4] These devices collect useful data with the help of various existing technologies and then automatically exchange data between other devices.^[5] Current market examples include smart thermostat systems and washer/dryers that utilize Wi-Fi for remote monitoring.

Besides the plethora of new application areas for Internet connected automation to expand into, IoT is also expected to generate large amounts of data from diverse locations that is aggregated very quickly, thereby increasing the need to better index, store and process such data.^{[7][8]}

So, we have a problem

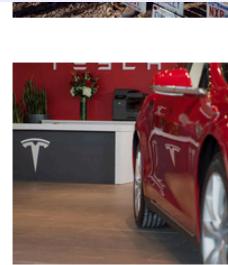
News articles can be
too ‘complex’
whereby readers might **leave**
the website and go to
competitor news websites.

Wow, did I miss something?!

What is the **Internet of Things**?

The purchase underlines how Nest, which Google bought for about \$3.2 billion earlier this year, is working to become a bigger player of connected devices for consumers. The trend is part of a technology movement dubbed the “**Internet of things**” everyday items are connected to gather data and be controlled by mobile devices.

The products gather data on consumer behavior. Privacy experts have questioned how Nest uses Nest’s data. Nest has said it is a privacy company.



*Wouldn't it be nice if **someone** could **explain this to you online**?*

Goal

To

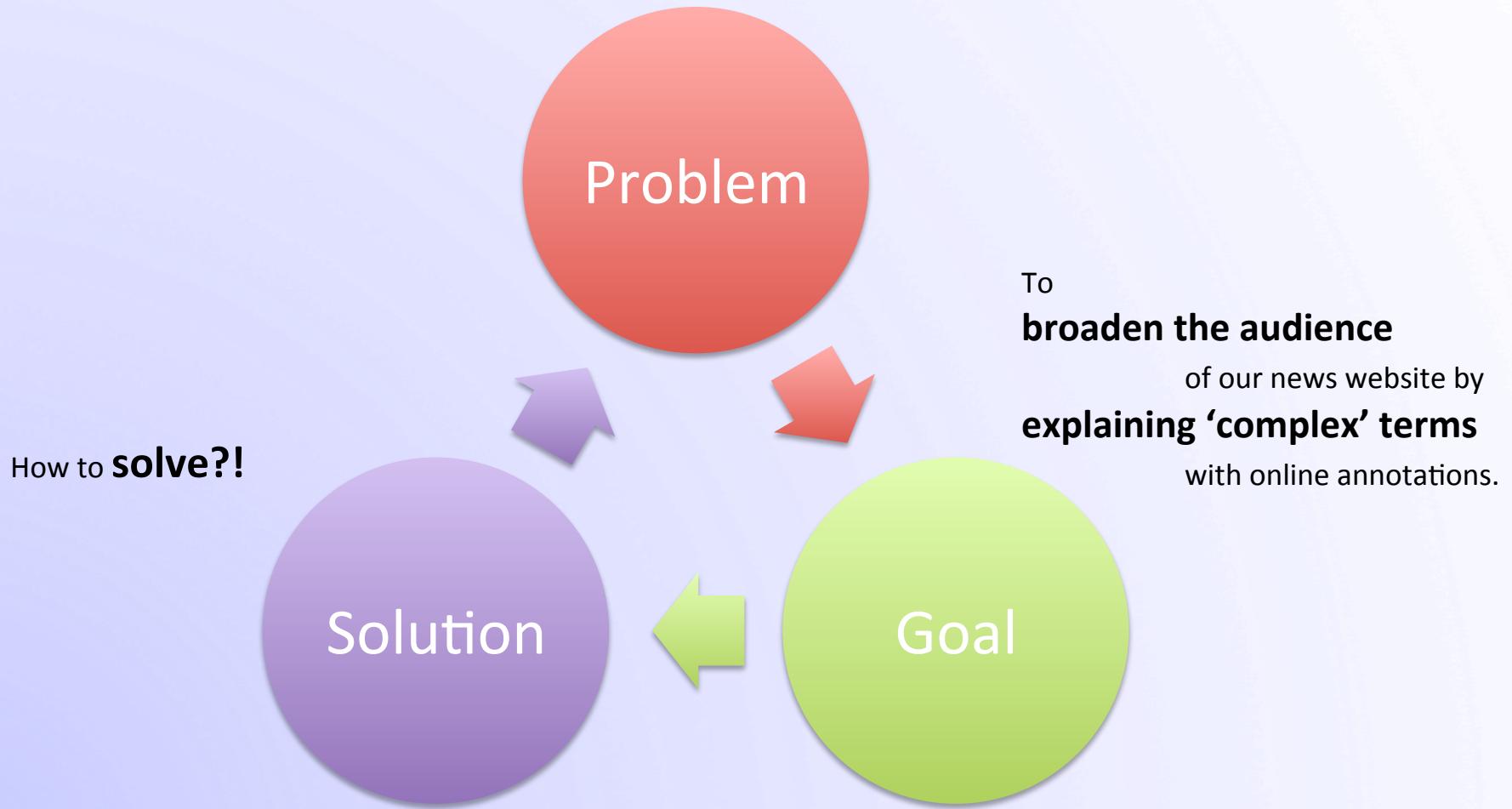
broaden the audience

of our news website by

explaining ‘complex’ terms

with online annotations.

News articles can be
too 'complex'
whereby readers might **leave**
the website and go to
competitor news websites.



'Explanation Retrieval'

- **Manual explanations**

Readers explain terms that other readers don't understand



Human Computation

- **Automatic explanations**

Retrieving related articles with semantic searching



Information Retrieval

But... Human Computation is scarce!!

Project Organization



Defining

Designing

Developing

Presenting

- Problem
- Goal
- Solution
- Scope
- Design
- Requirements
- IR and HC techniques
- MVP
- Testing
- Presenting
- Demo
- **Sell it!**

Design choices

Web Application

News articles

Back-end &
Database

Human Computation

IR Model

SKIIR-app



Scope

Chrome Extension

Bloomberg

Play Framework &
PostGreSQL

Readers

AlchemyAPI &
DBpedia &
New learnings

Features - What should it do?

0. **Scrape** news articles

1. **Request** explanation

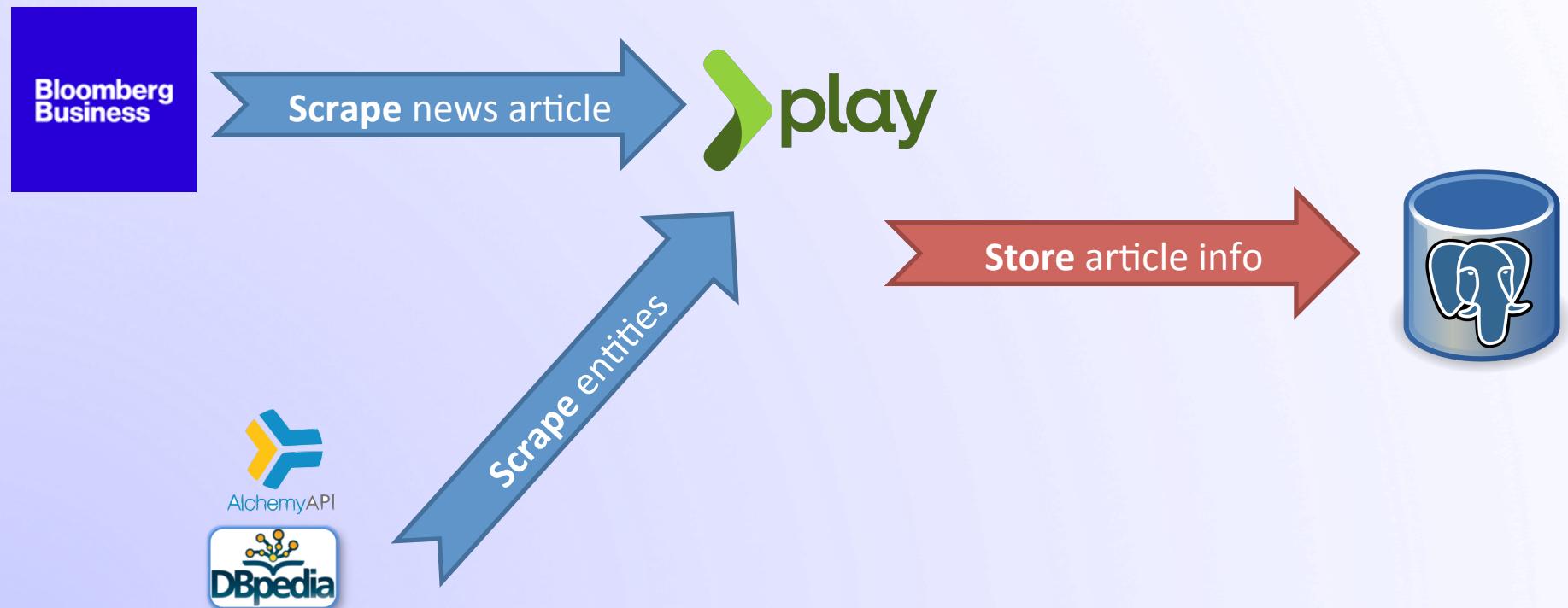
2. **Semantically** retrieve explanations

3. **Provide** explanation

4. **Review** explanation

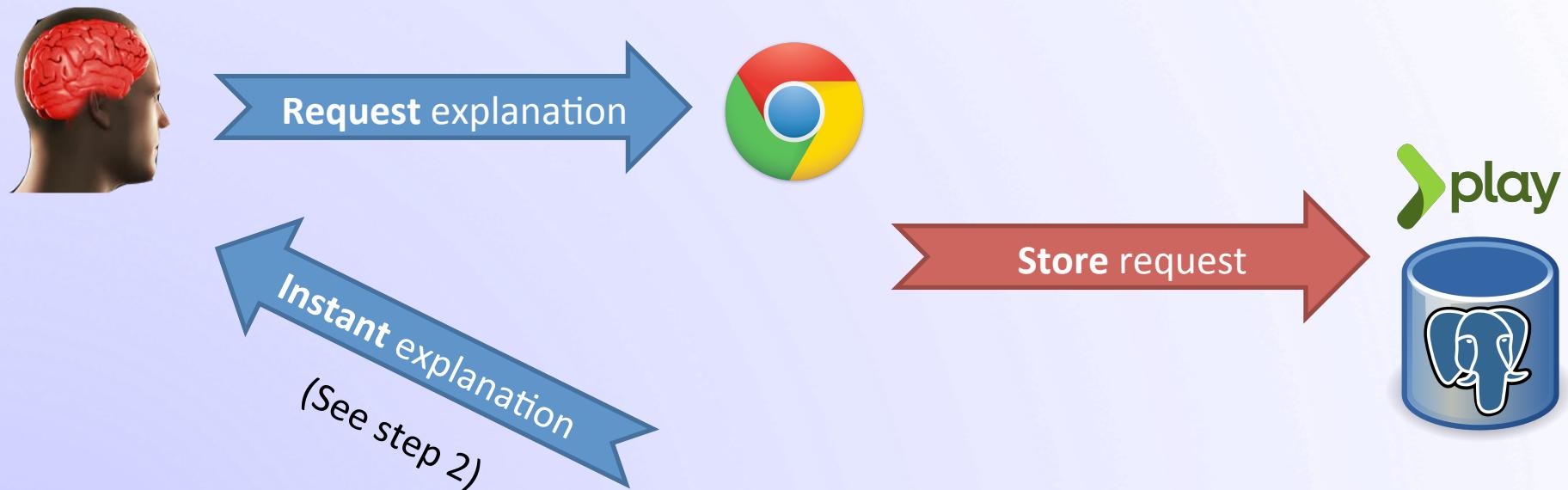
Workflow - How should it work?

0. Scrape news articles



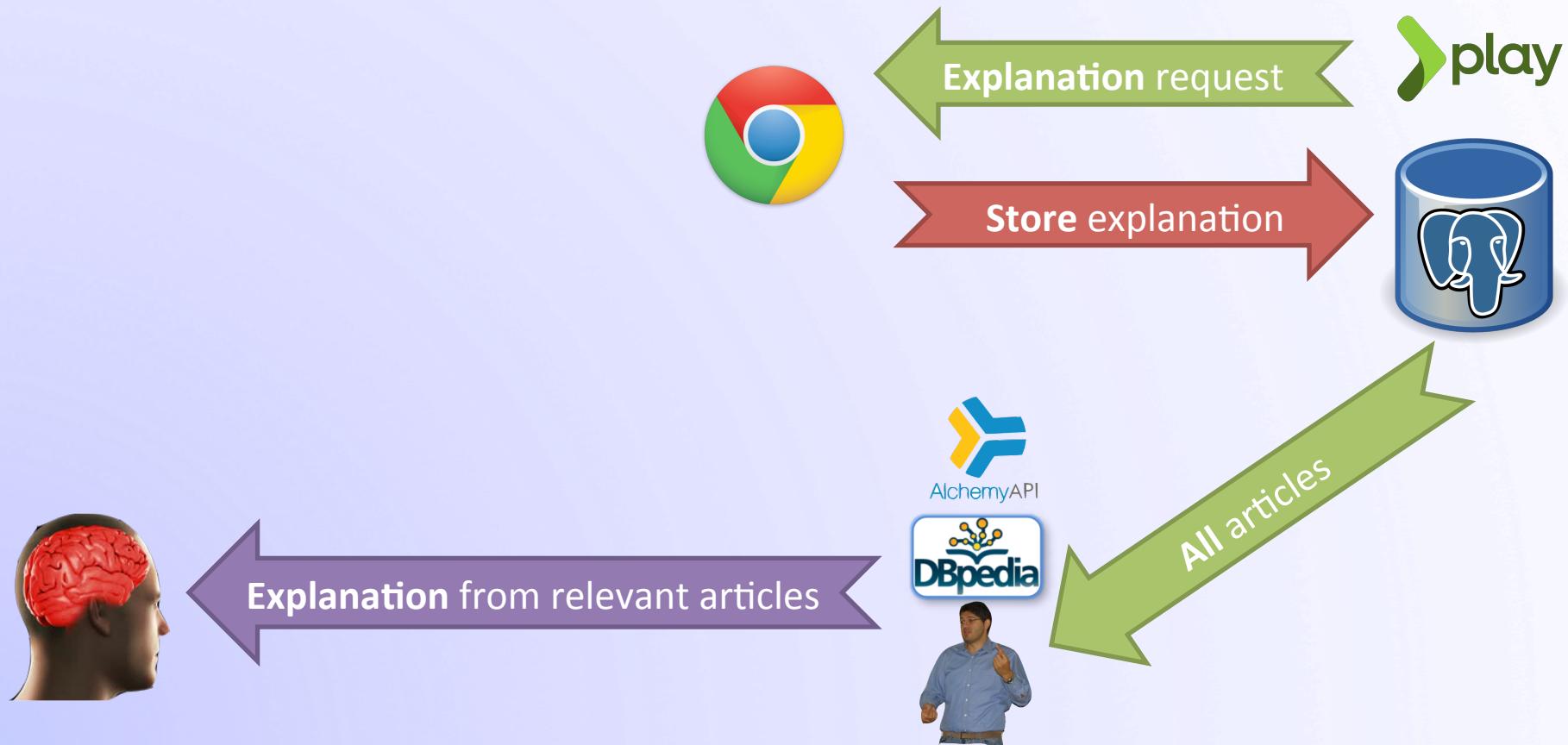
Workflow - How should it work?

1. Request explanation



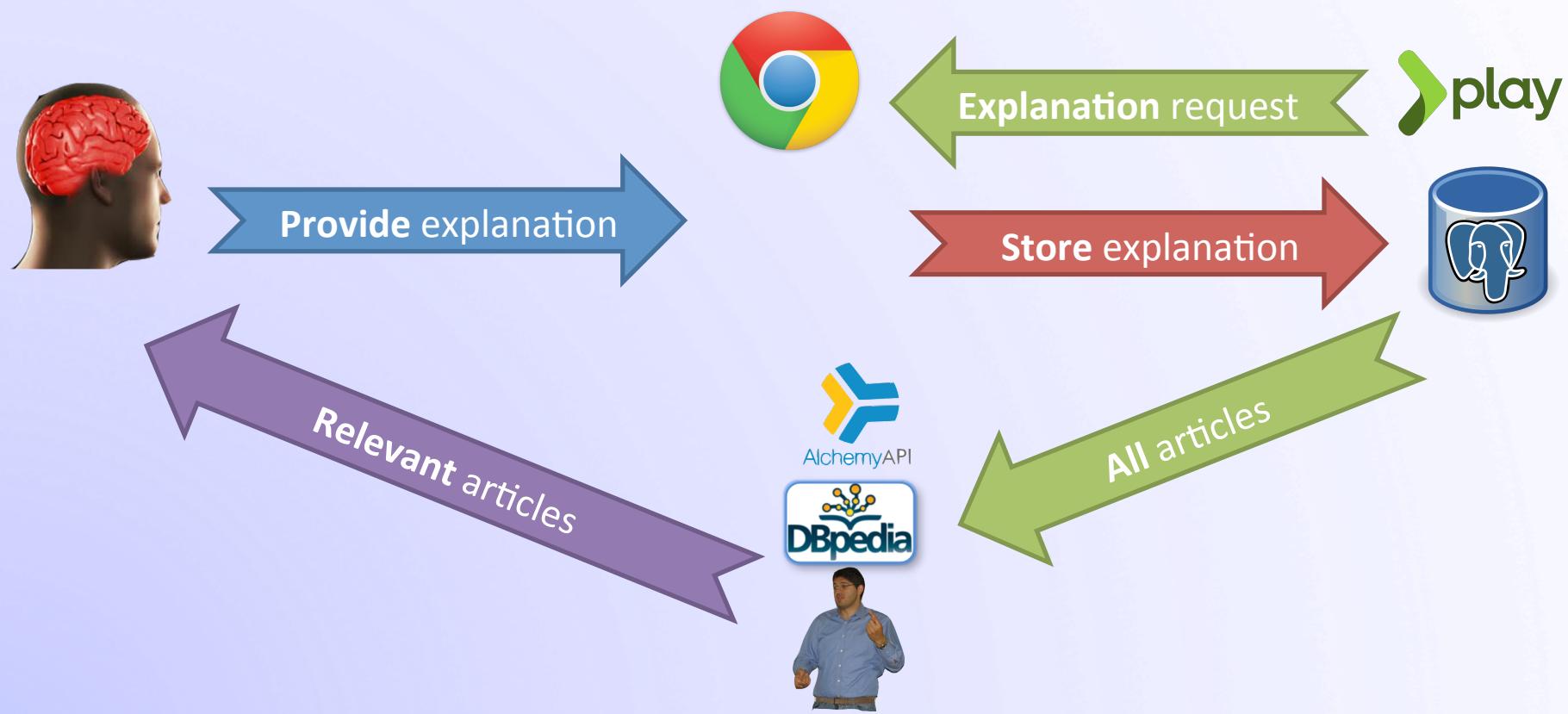
Workflow - How should it work?

2. Semantically retrieve explanation - automatically



Workflow - How should it work?

3. Provide explanation - manually

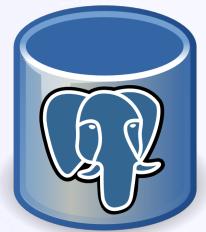
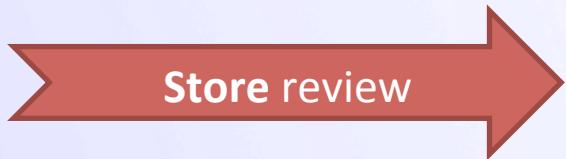
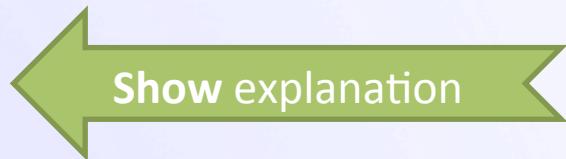


Workflow - How should it work?

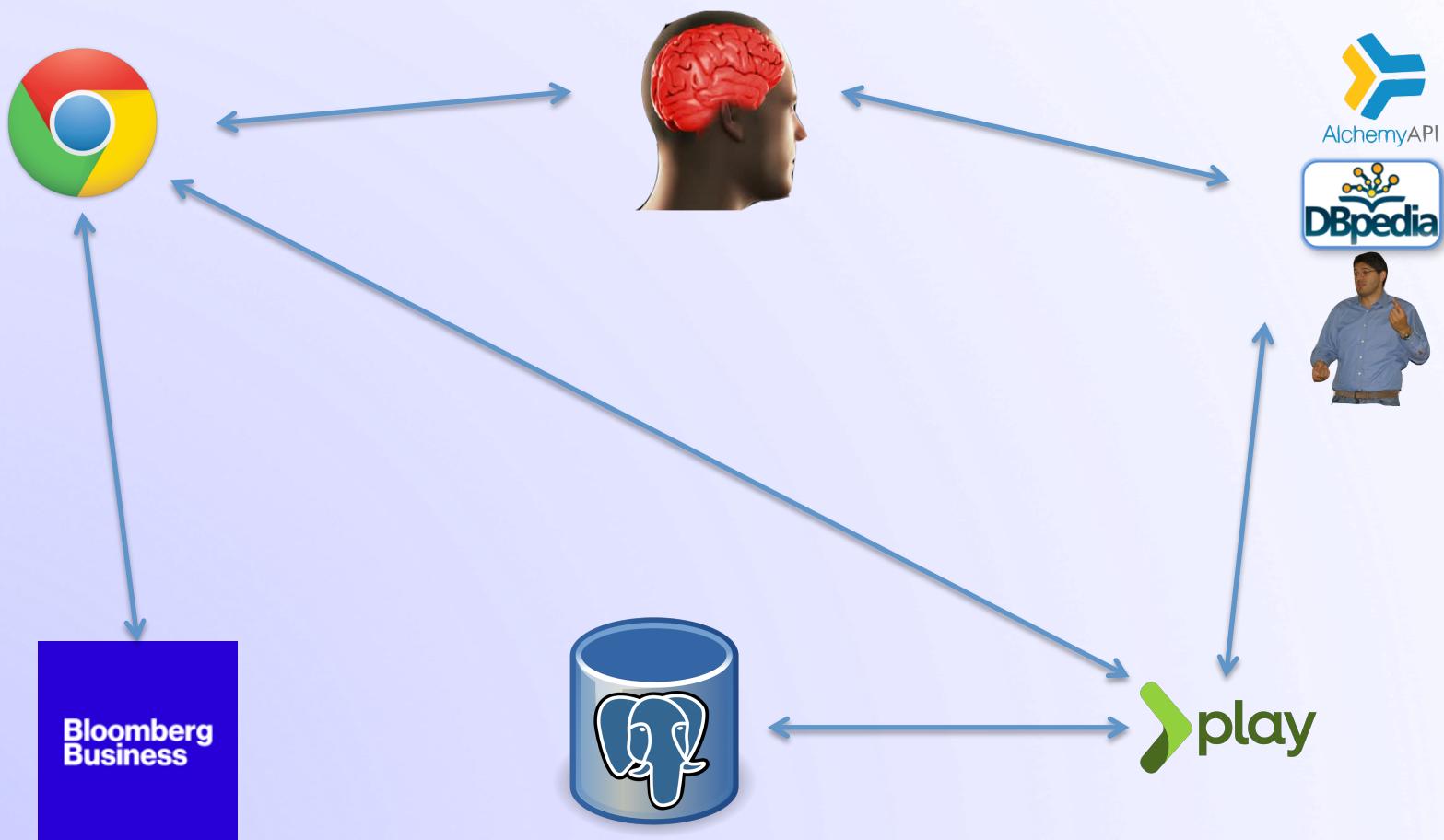
4. Review explanation



(see step 3)



Workflow – Functional overview



Explanation Retrieval

1. Extraction of requisites
2. Relevance of terms
3. SCS Connector
4. Adding diversity
5. Ranking results

Extraction of Requisites

- Entities
- Concepts
- Linked Data
- Relevance

Entity	Relevance	Sentiment	Type	Subtypes	Linked Data
Nest	0.840655	negative	Person		
Dropcam	0.362548	mixed	Person		
Google Inc.	0.356966	positive	Company	AcademicInstitution AwardPresentingOrganization OperatingSystemDeveloper ProgrammingLanguageDeveloper SoftwareDeveloper VentureFundedCompany	dbpedia freebase yago website
Nest Labs	0.311734	positive	Company		dbpedia freebase
Matt Rogers	0.311671	neutral	Person	MusicalArtist FootballPlayer TVPersonality TVActor	dbpedia freebase yago
Menlo Ventures	0.244707	positive	Company		dbpedia freebase
Dropcam Inc.	0.207606	positive	Company		

Concept	Relevance	Linked Data
Chief executive officer	0.91255	dbpedia freebase
Executive officer	0.840277	dbpedia freebase
Mountain View, California	0.784214	dbpedia freebase yago website
Silicon Valley	0.743894	dbpedia freebase yago
Facebook	0.691181	dbpedia freebase yago website
Yahoo!	0.689852	dbpedia freebase yago website

Ranking Results

Combining all to **rank** the articles on total **relevance**

$$S_B = \sum_i \{R_{iA} * R_{iB} * (1 - SCS_i)\}$$

Relevance of terms

Relevance of terms in a specific article

0,35

Petrobras creditors could request an acceleration of payments on some of its debt if the company doesn't release an audited financial statement by the end of 0,63

Public prosecutors are investigating 20 companies, including 16 builders, for participating in a alleged cartel to overcharge the oil company. OAS and Galvao Engenharia SA Brazilian construction companies, filed for bankruptcy after investigations started.

The president said the builders financial difficulties won't hold up planned tenders to build and operate infrastructure projects. She said she plans 0,11 Chinese investment for a railway linking Brazil's Atlantic coast with the Pacific Ocean

0,88
The Supreme Court has also authorized investigations of 49 politicians including Senate chief Renan Calheiros and lower house President Eduardo Cunha in relation to the alleged 0,78 kickbacks at Petrobras. Former official company turned state's witness said that hundreds of millions of dollars went to political parties, including Rousseff Workers' Party. The scandal has eroded the support she needs in Congress to approve austerity measures.

$$S_B = \sum_i \{ R_{iA} * R_{iB} * (1 - SCS_i) \}$$

SCS Connector

Relationship between entities

Petrobras creditors could request an acceleration of payments on some of its debt if the company doesn't release an audited financial statement by the end of May.

Public prosecutors are investigating 20 companies, including 16 builders, for participating in a alleged cartel to overcharge 0,23 company. OAS and Galvao Engenharia SA Brazilian construction companies, filed for bankruptcy after investigations started.

The president said the builders financial difficulties won't hold 0,56 hned tenders to build and operate infrastructure projects. She said she plans tap Chir 0,56 estment for a railway linking Brazil's Atlantic coast with the Pacific Ocean

The Supreme Court 0,76 so authorized invest 0 ions of 49 politicians including Senate chief Renan Calheiros and lower house President Eduardo Cunha in relation to the alleged kickbacks at Petrobras. Former officials at the company turned state's witness said that hundreds of millions of dollars went to political parties 0,83 ing Rousseff Workers' Party. The scandal has eroded the support she needs in Congress to approve austerity measures.

$$S_B = \sum_i \{ R_{iA} * R_{iB} * (1 - SCS_i) \}$$

Ranking Results

Combining all to **rank** the articles on total **relevance**

$$S_B = \sum_i \{ R_{iA} * R_{iB} * (1 - SCS_i) \}$$

S_B = Score of article B

R_{Ai} = Relevance of term i in article A

R_{Bi} = Relevance of term i in article B

SCS_i = SCS score of term i in article A

Adding Diversity

Maximal Marginal Relevance (MMR) method

Max Relevance

vs

Min Relevance

The company delayed in December reporting results for a second time as the board couldn't reach an agreement on the size of write-downs stemming from graft-related costs. Moody's Investors Service lowered Petrobras credit rating two levels to junk grade in February.

"I want to ensure that Petrobras goes back to the market," she said.

Petrobras creditors could force the company to negotiate some of its debt if the company doesn't release a audit report by the end of May.

Original document

Public prosecutors are investigating 20 companies, including 16 builders, for participating in a alleged cartel to overcharge the oil company OAS and Galvao Engenharia SA, Brazilian construction companies involved in the corruption investigations started.

The president said the builders financial difficulties won't hold up planned tenders to build and operate infrastructure projects. She said she plans tap Chinese investment for a railway linking Brazil's Atlantic coast with the Pacific Ocean.

The Supreme Court has also authorized investigations of 49 politicians including Senate chief Renan Calheiros and lower house President Eduardo Cunha in relation to the alleged kickbacks at Petrobras. Former officials at the company turned state's witness said that hundreds of millions of dollars went to political parties, including Rousseff Workers' Party. The scandal has eroded the support she needs in Congress to approve austerity measures.

Rousseff was rated good or very good by 23 percent of respondents, the lowest level she took office in 2011 and the worst for any Brazilian leader since Fernando Henrique Cardoso in 1999, according to a Datafolha poll published Feb. 7 on the website of newspaper Folha de S. Paulo. The rating tumbled from 42 percent in December, the pollster said.

Most poll respondents said they thought Rousseff was aware of corruption at Petrobras, it was happening, even though she has publicly denied such knowledge. Prosecutors police have alleged that company executives demanded bribes in return for approval. Standard & Poor's this month affirmed Brazil's investment grade status with a stable inflated construction contracts. The acquisition add to voter frustration with a stagnating economy and diminishing water supplies as the country's biggest city blazes. The S&P said the government's budget deficit will rise to 6.7 percent of gross domestic product.

"Petrobras is the main issue," Thiago da Cunha, a political consultant, said. "The stock market was c that the scar a whole."

Once a symbol of Brazil's rising econ plummeted this year to \$43 billion, d Investors Services and Fitch Ratings officials delayed plans to account for

While Petrobras stock recovered when Chief Executive Officer Maria das Gracas Foster stepped down Feb. 4, most gains were erased Feb. 6 when her replacement was named Almirin Bendine. Almirin Bendine was CEO of state-controlled bank Banco do Brasil and a ruling party collaborator.

To contact the reporter on this story:
jlevin20@bloomberg.net

The government posted a budget deficit of \$8.6 billion reais (\$16.6 billion) in February, compared with a median forecast for a gap of 2.8 billion reais in a Bloomberg survey.

Finance Minister Joaquim Levy has said spending cuts and tax increases are fundamental to avoid a credit downgrade. He has pledged to boost the primary surplus, which excludes interest payments, to 1.2 percent of gross domestic product this year from a 0.64 percent deficit in 2014.

With that as a target, the government has already capped some spending by ministries and is awaiting congressional approval for tax increases and cuts in social welfare benefits such as unemployment insurance and pensions.

Most poll respondents said they thought Rousseff was aware of corruption at Petrobras, it was happening, even though she has publicly denied such knowledge. Prosecutors police have alleged that company executives demanded bribes in return for approval. Standard & Poor's this month affirmed Brazil's investment grade status with a stable inflated construction contracts. The acquisition add to voter frustration with a stagnating economy and diminishing water supplies as the country's biggest city blazes. The S&P said the government's budget deficit will rise to 6.7 percent of gross domestic product.

"The correction will pave the way for stronger growth," S&P said, forecasting expansion of 2 percent for Latin America's largest economy next year following a 1 percent contraction in

ld

Other results

Demonstration

SKiIR

Explanation Retrieval

See the **application!**

Reflection - Accomplishments

SKiIR-app

Minimal Viable Product

- **Manual explanations**

Readers explain terms that other readers don't understand



Human Computation

- **Automatic explanations**

Retrieving related articles with semantic searching



Explanation Retrieval

Why should a newspaper use it?!

- **Broaden** the audience of your newspaper
- **Extend** visiting times
- **Expand** number of visitors

Why should a reader use it?!

- More **accessible** news articles
- More **background** information
- Reading **pay-walled articles** for free

How to incentivize readers?

- **Expert points**
Allocating expert points for (high) voted explanations
- **Pay-walled articles**
Granting free access to pay-walled articles
- **CrowdFlower**
Paying for explanations

Reflection

- **Manual explanations**

Readers explain terms that other readers don't understand



Human Computation

- Better explanations
- Supported by relevant articles
- Context
- Consistent (bad intention)
- Slower response
- Incentivizing is difficult

- **Automatic explanations**

Retrieving related articles with semantic searching



V S

Information Retrieval

- Instant response
- Consistent
- Based on relevant articles

- Context
- 'Computer-ish' explanations

Reflection – Difference with initial requirements

	Initial state	↔ Current state
Simplified IR	Complete ‘explanation’	↔ Sentences from relevant articles
Measuring success	Usage statistics	↔ Testing software functionality
Integration	CrowdFlower	↔ Independent browser extension
Expert incentivizing	Reward expert points	↔ No expert incentivizing

Reflection - Problems

Scope

- Defining exact requirements of the design

Design

- Not relying only on Human Computation

Teamwork

- Integration of different system parts
- Communication

Minimal Viable Product

- Sophisticating the Information Retrieval model

Reflection – Future work and research

- **Improve** semantic search model
 - Learning from the great ones (Google News search)
- **Define** business model
 - How to enter the market?
 - (Sell it to newspapers or publish extension?)
- **Incentivizing** readers to explain
 - Earning ‘expert status’
 - Access to pay-walled articles
 - More measurements?

Bibliography

1. **Bozzon.** (2015). Syllabus IN4325. TU Delft.
2. **Ceri, Bozzon, & etAlumni.** (2013). Web Information Retrieval. Springer.
3. **Law, & Ahn, v. (2011).** Human Computation. Morgan & Claypool.
4. **Nunes & et Alumni.** (2014). SCS Connector – Quantifying and Visualising Semantic Paths Between Entity Pairs.