NL search for semantic web

Lukas Kleine Büning
Pichaya Kanjanapisith
Yuchun Chen
Venkat

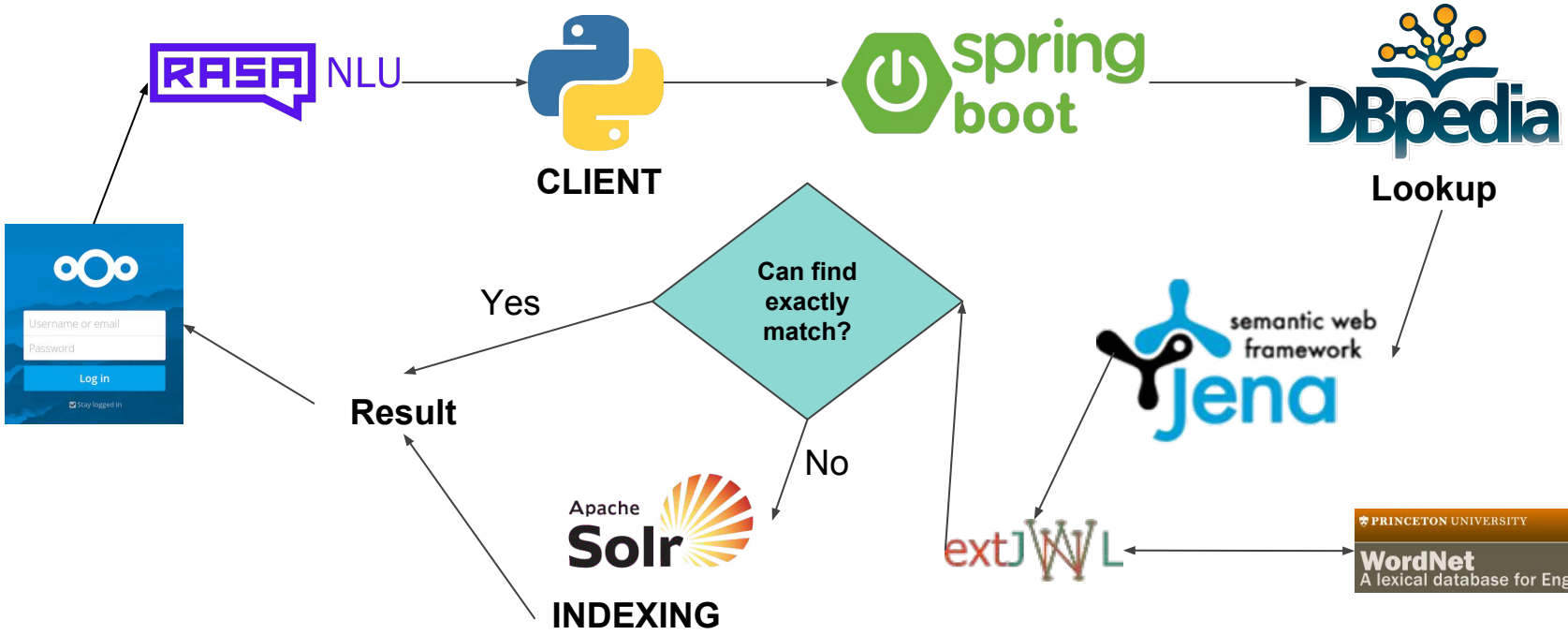
Agenda

- Motivation & Goals
- Component
- Demo
- Outcome
- Future work
- Responsibility

Motivation & Goals

- **Motivation:**
  - Searching in semantic sources requires special knowledge (SPARQL, ...)
  - General users cannot gain any benefit from such data sources
- **Goals**:
  - Creation of an interface which translates natural language searches into semantic web queries to run against data pools like dbpedia.org

Components : Application flow



Components

RASA for NLU processing

- As the last presentation.

Components

CLIENT

Python client for REST consumer

- Simple REST consumer, make api call to the backend with the input from RASA.

Spring boot server.

- Provide a simple web service which is interface connect between java backend and Rasa NLU.

Components

DBpedia Lookup : Subject measurement

- Web service that can be used to look up DBpedia URIs by related keywords.
- The results are ranked by the number of inlinks pointing.
- For example,
  - Subject input is “USA” will output as United_States
  - Subject input is “Berl” will output as Berlin.

Components

Apache Jena - SPAQL property query

- Open source framework for OWL and Semantic web.
- Use for query all property by the given Subject.
- The backend will find the property that match exactly with predicate then output as result.

.

Components

WordNet - Dictionary dataset for predicate

- English similarity dictionary linked by for synonym and generality of the word call synset.
- Synsets are interlinked by means of conceptual-semantic and lexical relations.
- For example:
    - {furniture, piece_of_furniture} the similarity will be increase to specific ones like {bed}.

.

Components

extJWNL - Predicate measurement

- Java API for creating, reading and updating dictionaries over WordNet.
- The key feature of this framework is easy to extent the dictionary dataset and possible to link it to the other data source.
- This procedure will receive predicate from RASA trying to find similarity word from the dictionary and output all possible word.
- For example:
    - It is possible to add dateOfBirth as the child of date of birth entity so when user input “What is the date of birth of Donald Trump?” it will be output dateOfBirth as one of the predicate.

Components

Apache Solr - Indexing framework

- Solr is a standalone enterprise search server with a REST-like API.
- Solr index created once, can be used for fast look up of relavant documents.
- Solr framework also offers hit highlighting,spellcheck, auto-suggest features on the built index.

Demo

Outcome

- User could search against data pool with input query.
- Even the input query does not match with training data, we provide data index to generate the result
- Gain knowledge about semantic search, natural language processing and page indexing

Future work

- The extension
  - Front-ends
    - Pagination
    - Question suggestion based on NLU training datasets
  - RASA NLU
    - Increase training datasets
  - ExtJWNL dictionary
    - Increase dictionary coverage
  - Solr indexing
    - Increase indexing pages

Responsibilities

- Lukas Kleine Büning
  - RASA NLU, Python Client
- Pichaya Kanjanapisith
  - Spring boots Server, DBpedia lookup, extJWNL , Jena Query
- Yuchun Chen
  - Webinterface , Python Client, Solr Indexing
- Venkat
  - Solr Indexing , evaluation

Thank you!

Q&A

Repository :
https://gitlab.tubit.tu-berlin.de/pkanjan37/SW-LD_NLP_project

Reference

https://discuss.elastic.co/t/semantic-search-engine-on-the-top-of-es-any-suggestions-comments/41527

http://lucene.apache.org/solr/4_6_1/

http://mudassirshahzad.com/wp-content/uploads/2017/02/spring-boot.png

https://pbs.twimg.com/profile_images/578479910366269440/quS6q6Yu.png