GERBIL

A BENCHMARKING NAMED ENTITY RECOGNITION AND LINKING CONSISTENTLY
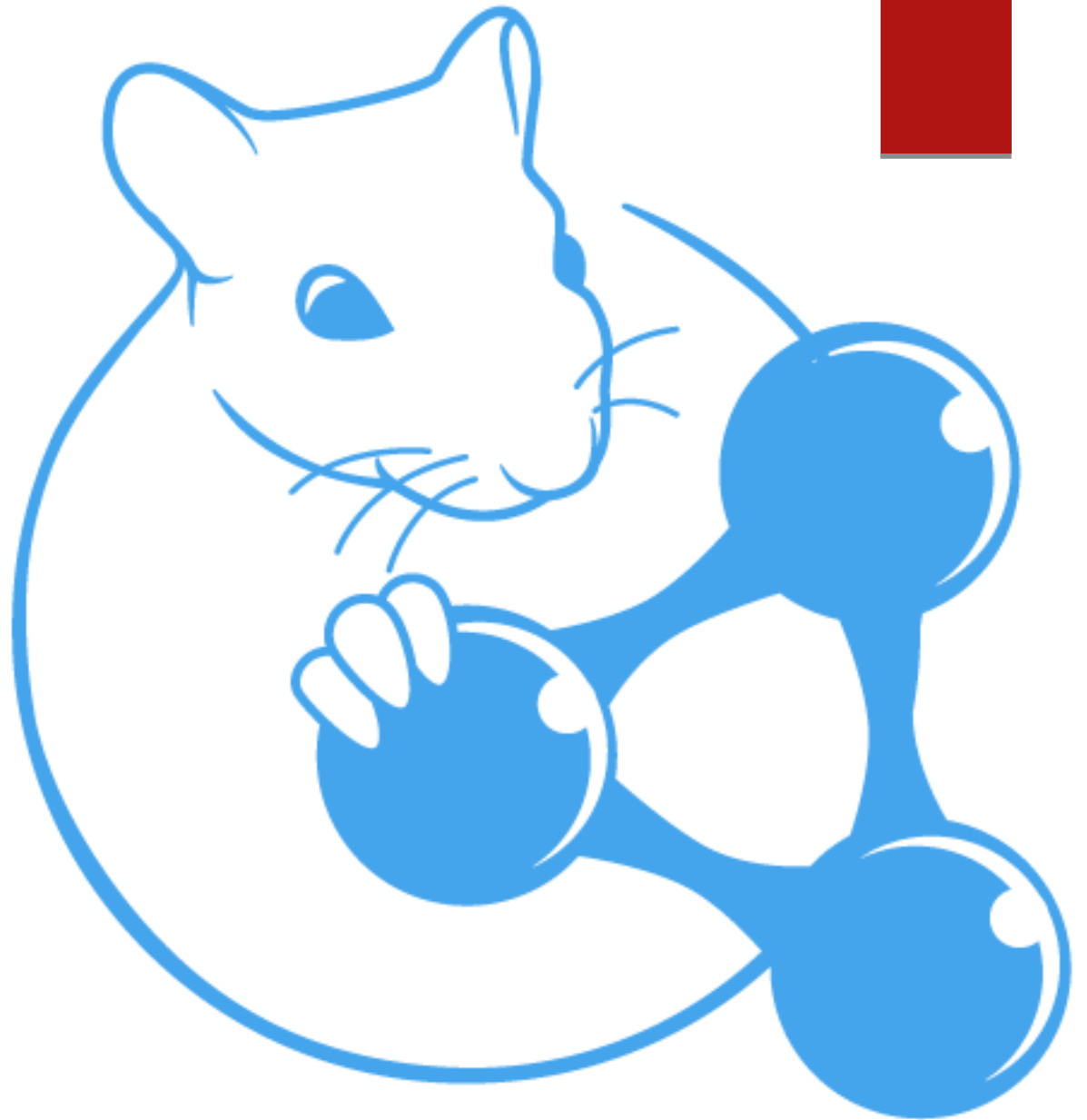
Table of Contents

ANNOTATORS

- Structured data and Un-structured data.

- The need to bridge between the unstructured data on the document Web and the structured data on the Data Web has led to the development of a considerable number of annotation tools.

- Example of annotated block of text:

  Jim bought 300 shares of Acme Corp. in 2006.

- An annotated block of text that highlights the names of entities:

  [Jim]Person bought 300 shares of [Acme Corp.]Organization in [2006]Time.

- Examples-Tagme,FREME,FOX etc

ANNOTATORS Contd.

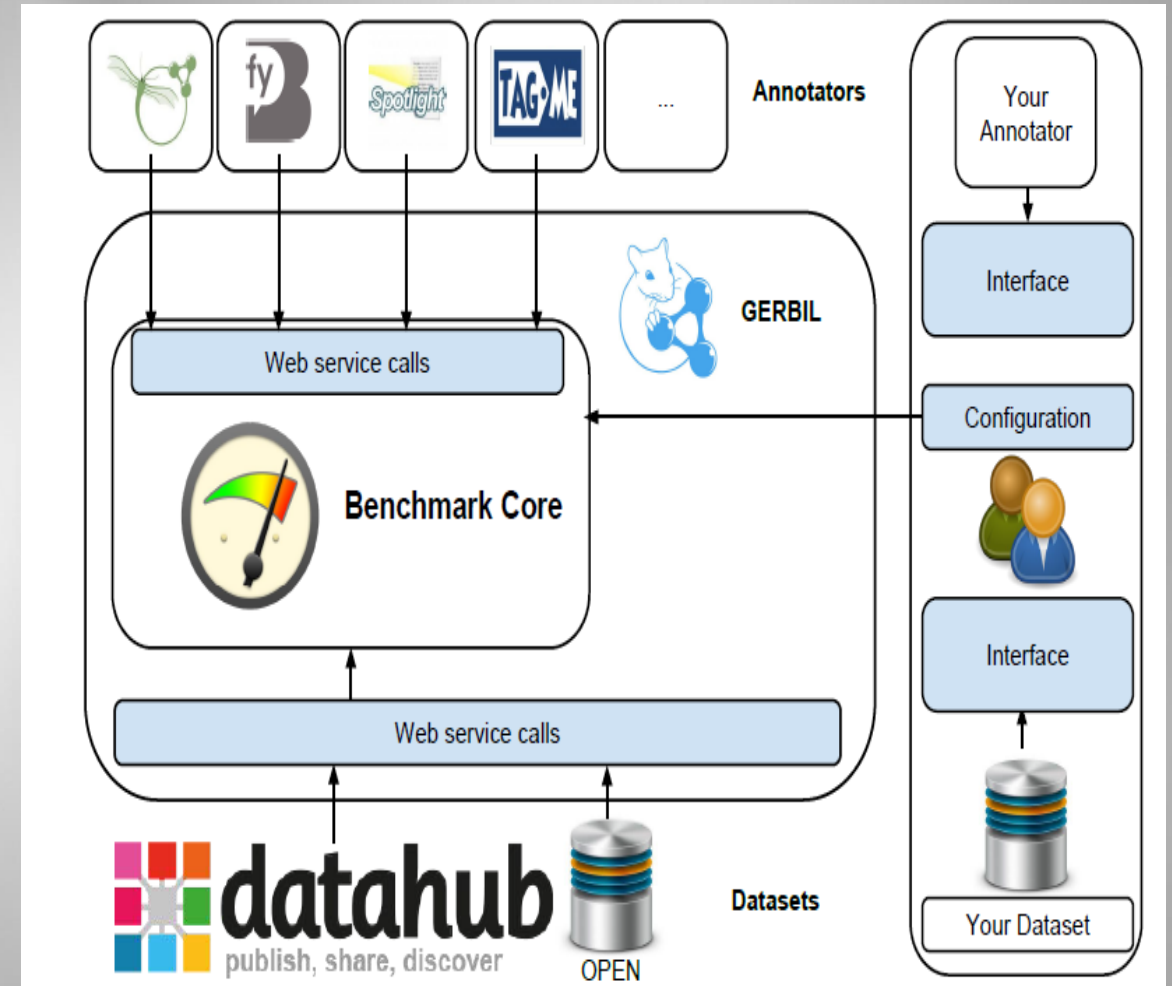- In the domains of named entity recognition and entity linking, the large number of systems and their orthogonal evaluation w.r.t. measures and datasets has led to an unclear landscape pertaining to the abilities and weaknesses of the different frameworks.

- These tools are hard to compare since published results are calculated on diverse datasets and measured in different units.

GERBIL - INTRODUCTION

▶ GERBIL—a general entity annotation system based on the BAT-Framework.

▶ BAT frame work- A framework to compare text annotators: systems that, given a text document, aim at finding the entities the text is about, identified as Wikipedia pages.

▶ GERBIL offers an easy-to-use web-based platform for the agile comparison of annotators using multiple datasets and uniform measuring approaches.

▶ The framework generates data in machine-processable format, allowing for the efficient querying and post-processing of evaluation results.

GERBIL-Architecture

- A service-oriented architecture driven by the model-view-controller pattern.

- Open Source framework.

- The output of experiments as well as descriptions of the various components are stored in a server less database for fast deployment.

- To add a tool to GERBIL, all the end user has to do is to provide a URL to a REST interface to its tool which abides by a given specification.

- The integration and benchmarking of the tool against user-specified datasets is then carried out automatically by the GERBIL platform.

- Currently 20 different annotators on 12 different datasets within 6 different experiment types.

GEBRIL-List of annotators

- Current annotator tools will search the Wikipedia/DBPedia for the user given input text and results the corresponding page from them.

- In this case if any information provided in the text is not in Wiki they will not be listed in the result set.

- Unlike above this framework can deal with several annotators, data sets and link to any knowledge base.

	BAT-FRAMEWORK	GERBIL 1.0.0	GERBIL 2.5
Wikipedia Miner	✓	✓	(✓)
Illinois Wikifier	✓	(✓)	✓
Spotlight	✓	✓	✓
AIDA	✓	✓	✓
TagMe 2	✓	✓	✓
NERD-ML		✓	✓
KEA		✓	✓
WAT		✓	✓
Dexter		✓	✓
AGDISTIS	(✓)	✓	✓
Babelfy		✓	✓
FOX			✓
FRED			✓
FREME			✓
entityclassifier.eu			✓
CETUS			✓
xLisa			✓

Advantages of GERBIL

- Easy integration of annotators
  - Provides a wrapping interface that allows annotators to be evaluated via their HTTP interface.
- Easy integration of datasets
  - Provides means to gather datasets for evaluation directly from data services such as DataHub.
- Easy addition of new measures
  - The evaluation measures used by GERBIL are implemented as interfaces. Thus, the framework can be easily extended
- Extensibility
  - GERBIL is provided as an opensource that can be extended by members of the community both to new tasks and different purposes.

Thank you!!