



On the Importance of Drill-Down Analysis
for Assessing Gold Standards and Named Entity Linking Performance
Fabian Odoni, **Philipp Kuntschik**, Adrian Brasoveanu, Albert Weichselbraun

Annotating Entities

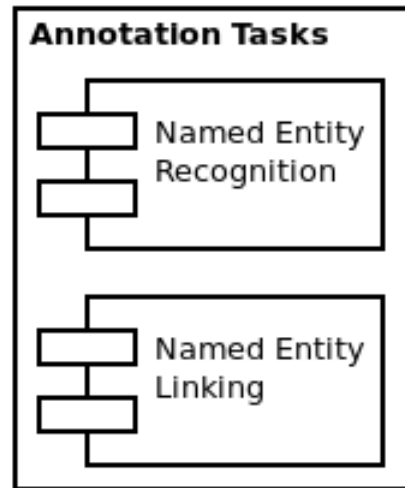
Unknown Document

Clinton trolls Trump ahead of Putin summit with World Cup tweet

"Great World Cup. Question for President Trump as he meets Putin: Do you know which team you play for?" Clinton tweeted Sunday night, ahead of the President's first official summit with Putin, taking place in Helsinki, Finland.

Earlier Sunday, Trump tweeted out congratulations to the tournament's victors, France, as well as Putin and his countrymen for "putting on a truly great World Cup Tournament -- one of the best ever!"

Clinton followed up on her jab at the President on Monday evening following the summit, when she wrote: "Well, now we know."

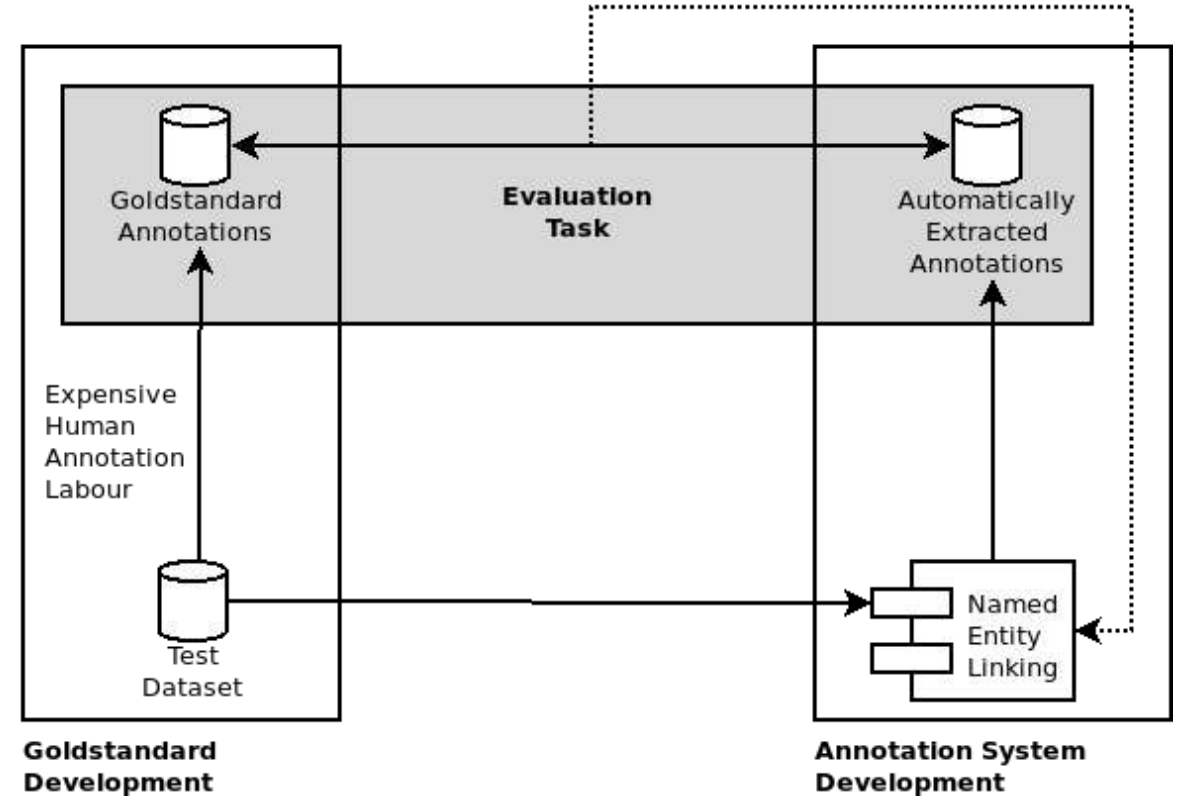


Annotations

```
dbpedia.org/resource/Hillary_Clinton
* "Clinton"
* "Clinton"
* "Clinton"
dbpedia.org/resource/Donald_Trump
* "Trump"
* "President Trump"
* "President"
* "Trump"
* "President"
dbpedia.org/resource/Vladimir_Putin
* "Putin"
* "Putin"
* "Putin"
* "Putin"
dbpedia.org/resource/FIFA_World_Cup
* "World Cup"
* "World Cup"
* "World Cup Tournament"
dbpedia.org/resource/Helsinki
* "Helsinki, Finland"
dbpedia.org/resource/France
* "France"
```

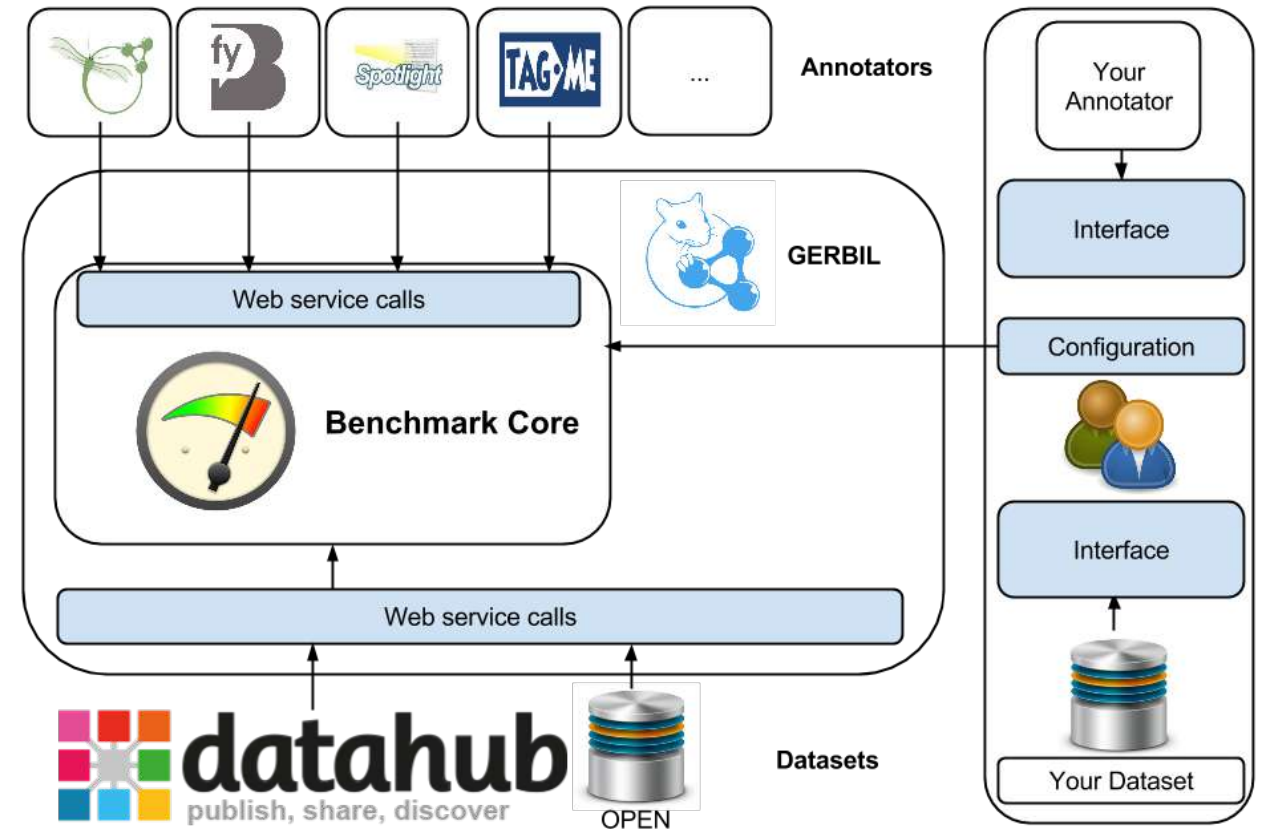
Annotation Service Development Cycle and the Role of Evaluations

- An annotation systems performance is measured by comparing the returned annotations to a goldstandard.
- Goldstandard development and annotation system development are separate, but not independent processes.
- (The evaluation results are used to improve the annotation system.)



Related Work: GERBIL (General Entity Annotation Benchmark Framework)

- An easy-to-use web-based platform for the agile comparison of annotators using multiple datasets and uniform measuring approaches.
- Used to compare annotators in NEL challenges (e.g. ESWC Open Knowledge Extraction Challenge)



Related Work: GERBIL (General Entity Annotation Benchmark Framework)

GERBIL Experiment

Experiment URI: <http://gerbil.aksw.org/gerbil/experiment?id=201801280015> and <http://w3id.org/gerbil/experiment?id=201801280015>

Type: A2KB

Matching: Mw - weak annotation match

Annotator	Dataset	Micro			Macro			InKB		InKB		EE		EE			
		F1 score	Micro Precision	Micro Recall	F1 score	Macro Precision	Macro Recall	F1 score	Macro Precision	Macro Recall	F1 score	Macro Precision	Macro Recall	F1 score	Macro Precision	Macro Recall	
AIDA	KORE50	0.584	0.6887	0.5069	0.5258	0.621	0.4747	0.5285	0.621	0.4813	0.5863	0.6887	0.5105	0.98	0.98	0.98	0
AIDA	KORE50 Entity Recognition	0.808	0.9528	0.7014	0.7559	0.8947	0.6913										
AIDA	KORE50 D2KB	0.575	0.7188	0.4792	0.4992	0.5977	0.4447	0.5018	0.5977	0.4513	0.5774	0.7188	0.4825	0.98	0.98	0.98	0
Babely	KORE50	0.5645	0.5664	0.5625	0.5263	0.547	0.548	0.5303	0.547	0.5547	0.5664	0.5664	0.5664	0	0	0	0
Babely	KORE50 Entity Recognition	0.726	0.7162	0.7361	0.6991	0.7232	0.7287										
Babely	KORE50 D2KB	0.6639	0.81	0.5625	0.5915	0.6803	0.5453	0.5955	0.6803	0.552	0.6667	0.81	0.5664	0.98	0.98	0.98	0

Related Work: TAC KBP nel_eval

- Console based implementation of evaluation measures for whole-document named entity linking evaluation.
- Result as CSV:

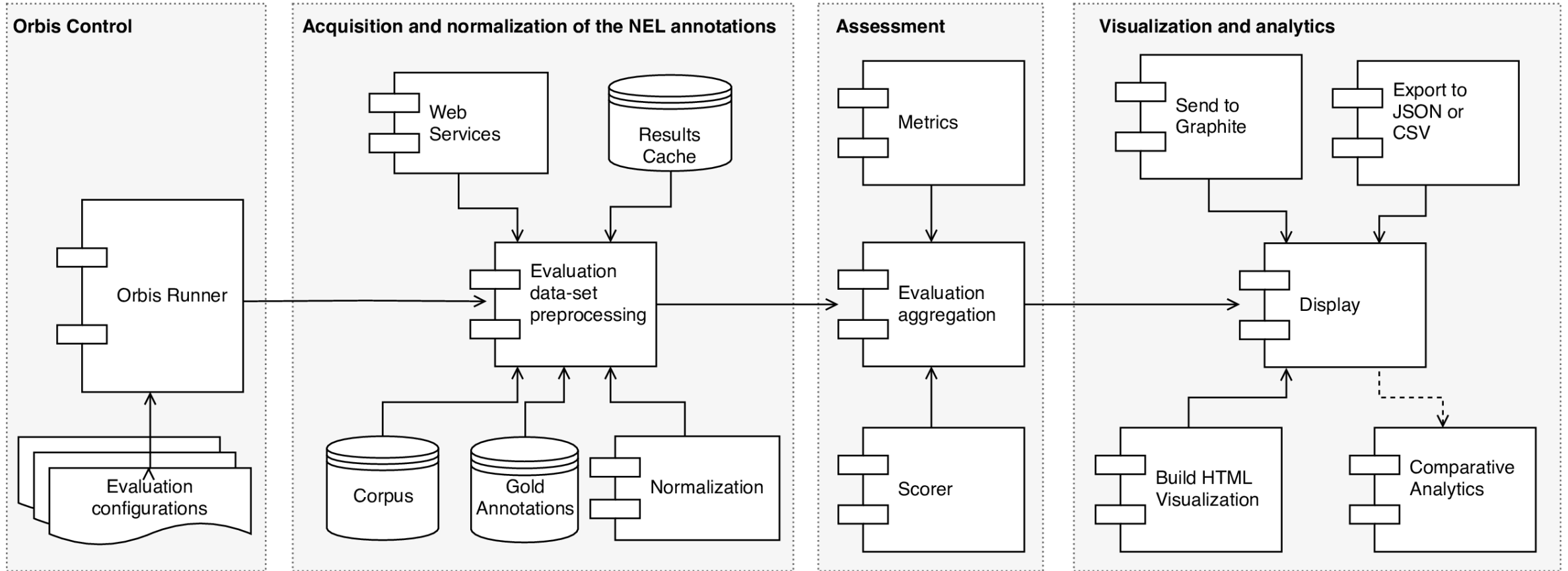
output	document	start	end	goldstandard	system
correct link	60	298	323	g"http://dbpedia.org/resource/Luxembourg_Stock_Exchange"	s"http://dbpedia.org/resource/Luxembourg_Stock_Exchange"
extra	60	164	179		s"http://dbpedia.org/resource/Capital_Markets_Authority_(Uganda)"
extra	60	492	496		s"http://dbpedia.org/resource/1987_(artist)"
extra	60	150	163		s"http://dbpedia.org/resource/Merrill_Lynch"
extra	60	369	372		s"http://dbpedia.org/resource/American_Automobile_Association"
extra	60	325	332		s"http://dbpedia.org/resource/Eastman,_Georgia"
extra	60	0	13		s"http://dbpedia.org/resource/Eastman_Kodak"
missing	60	0	16	g"http://dbpedia.org/resource/Eastman_Kodak"	
missing	60	150	179	g"http://dbpedia.org/resource/Merrill_Lynch"	
wrong-link	61	677	685	g"http://dbpedia.org/resource/Getty_Oil"	s"http://dbpedia.org/resource/Pennzoil"
correct link	61	190	198	g"http://dbpedia.org/resource/Pennzoil"	s"http://dbpedia.org/resource/Pennzoil"
extra	61	655	664		s"http://dbpedia.org/resource/Getty_Oil"
extra	61	513	518		s"http://dbpedia.org/resource/Texas"
extra	61	0	12		s"http://dbpedia.org/resource/First_Boston"
extra	61	69	75		s"http://dbpedia.org/resource/Texaco"
extra	61	358	363		s"http://dbpedia.org/resource/Court_(District_Electoral_Area)"

Introducing: ORBIS

- A versatile framework for annotation task evaluations.
- Its main purpose is to aid experts in detecting strenghts and weaknesses of their system.
- Visualizes gold standard and annotator results within their context (the document).
- Provides links to the used knowledgebase

- Main components:
 - Orbis Runner: Orchestrates the pipeline, configured with YAML files
 - Evaluation Pipeline: aquires, assesses the data; visualizes the results.

Orbis: System Architecture



Aggregator Name: recognize
Aggregator Profile: semantics
Aggregator Limit: None
Aggregator Service: recognize

Aggregator Dataset: reuters128
Evaluator Name: NEL_Evaluator
Scorer Name: NEL_ScorerOverlap
Entities: Person, Organization, Place

Some Score: 97
No Score: 31
Empty Responses: 12

Precision: 0.576
Recall: 0.571
F1 Score: 0.574

Stats for Item 106

Precision: 0.272
Recall: 0.750
F1 Score: 0.399

Gold

Plans to rejuvenate Polands economy by reducing central government control would help reassure Western creditors that the countrys economy was safe to invest in, a senior Polish official said. The business of granting loans to [Poland](#) is not as bad a business as you might imagine, senior Polish government spokesman Jerzy Urban told a news conference in [Stockholm](#). Urban, visiting the Swedish capital to deliver a lecture at the Foreign Policy Institute, announced earlier this week that [Poland](#) would soon offer shares to private citizens in state companies in a bid to make the economy more responsive. This was part of a major economic reform to be announced in the coming weeks, he said. Urban said the main problem with his countrys foreign debt burden of 32 billion dollars was short term interest charges but the long term looked more secure. He said he hoped talks under way with the [Paris Club](#), grouping Polands main government creditors, would shortly solve the problem of short term servicing. Urban told reporters that debt servicing was hampering the expansion of the national economy, but that the new plans for economic reform would open new vistas for Polish exports. He criticised the blocking efforts of some Western governments, whom he declined to name, for putting political considerations ahead of economic co-operation. After the lifting of U.S. Sanctions against the Polish economy we considered all political obstacles had been removed, but that has not been the case, Urban said.

Gold Entities

[Poland](#) (<http://dbpedia.org/resource/Poland>): 227 - 233

[Stockholm](#) (<http://dbpedia.org/resource/Stockholm>): 354 - 363

[Poland](#) (<http://dbpedia.org/resource/Poland>): 488 - 494

[Paris Club](#) (http://dbpedia.org/resource/Paris_Club): 891 - 901

True Positives: 3
False Positives: 8
False Negatives: 1

Computed

Plans to rejuvenate Polands economy by reducing central government control would help reassure Western creditors that the countrys economy was safe to invest in, a senior Polish official said. The business of granting loans to [Poland](#) is not as bad a business as you might imagine, senior Polish government spokesman [Jerzy Urban](#) told a news conference in [Stockholm](#). [Urban](#), visiting the Swedish capital to deliver a lecture at the [Foreign Policy Institute](#), announced earlier this week that [Poland](#) would soon offer shares to private citizens in state companies in a bid to make the economy more responsive. This was part of a major economic reform to be announced in the coming weeks, he said. [Urban](#) said the main problem with his countrys foreign debt burden of 32 billion dollars was short term interest charges but the long term looked more secure. He said he hoped talks under way with the [Paris Club](#), grouping Polands main government creditors, would shortly solve the problem of short term servicing. [Urban](#) told reporters that debt servicing was hampering the expansion of the national economy, but that the new plans for economic reform would open new vistas for Polish exports. He criticised the blocking efforts of some Western governments, whom he declined to name, for putting political considerations ahead of economic co-operation. After the lifting of [U.S.](#) Sanctions against the Polish economy we considered all political obstacles had been removed, but that has not been the case, [Urban](#) said.

Computed Entities

[Poland](#) (<http://dbpedia.org/resource/Poland>): 227 - 233

[Jerzy Urban](#) (http://dbpedia.org/resource/Jerzy_Urban): 316 - 327

[Stockholm](#) (<http://dbpedia.org/resource/Stockholm>): 354 - 363

[Urban](#) (http://dbpedia.org/resource/Jerzy_Urban): 365 - 370

[Foreign Policy Institute](#) (http://dbpedia.org/resource/Foreign_Policy_Institute): 429 - 453

[Poland](#) (<http://dbpedia.org/resource/Poland>): 488 - 494

[Urban](#) (http://dbpedia.org/resource/Jerzy_Urban): 691 - 696

[Paris](#) (<http://dbpedia.org/resource/Paris>): 891 - 901

[Urban](#) (http://dbpedia.org/resource/Jerzy_Urban): 1004 - 1009

[U.S](#) (http://dbpedia.org/resource/United_States): 1363 - 1366

[Urban](#) (http://dbpedia.org/resource/Jerzy_Urban): 1493 - 1498

Conclusion

- Introduced Orbis as an **annotation evaluation framework**.
- Focuses on **explaining annotation errors**, not on benchmarking.
- Demonstrated how Orbis helps detecting problems
 - in the annotators results.
 - in the goldstandard.
- Orbis will be available on Github in 2018.

Thank you very much for your attention.

Philipp.Kuntschik@htwchur.ch