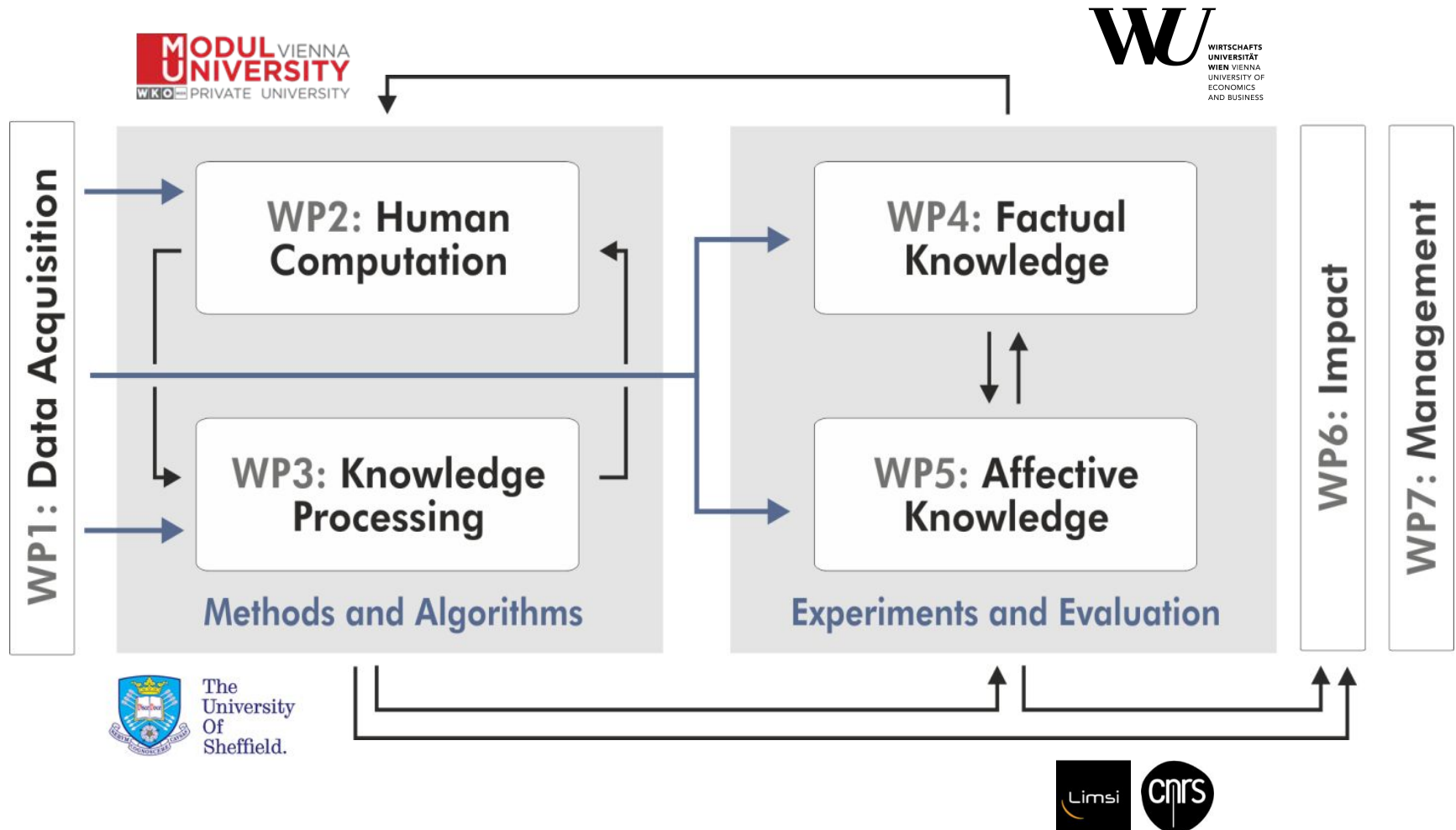


Embedded Human Computation for Knowledge Extraction and Evaluation

- **University of Sheffield**
Department of Computer Science
HC for NLP, GATE Text Mining Toolkit
- **MODUL University Vienna**
Department of New Media Technology
Games with a Purpose, Semantic Technologies
- **Vienna University of Economics and Business**
Research Institute for Computational Methods
Factual Knowledge Extraction, Ontology Learning
- **LIMSI-CNRS**
Man-Machine Communication Department
Affective Knowledge Extraction, Evaluation



Project Overview



- **Extensible Web Retrieval Toolkit (eWRT)**
Open Source Library
www.weblyzard.com/ewrt



- **Media Watch on Climate Change**
Multilingual Content Repository
www.ecoresearch.net/climate

- **Data Sources**

News and Social Media, Web Sites of Companies and Environmental Organizations

- **Data Volume**

10 Million Documents per Month

- **Multilinguality**

English, German, French, Spanish



EN, FR, DE, ES

Media Watch on CLIMATE CHANGE

[About](#) | [Help](#) | [Login](#)

Q solar

Documents

0-250 out of 8191

Date Range

27 Feb 2016 - 27 Apr 2016

Data Sources

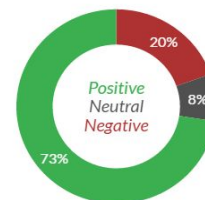
News / All

Interface Options

Unfiltered / EN

Visualizations

Export

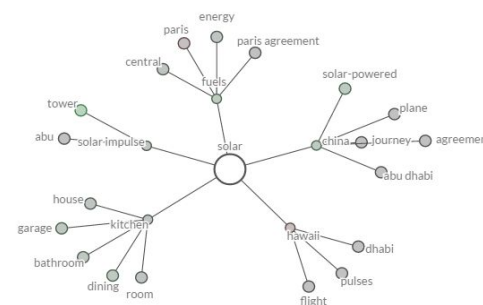


solar

- in the
 - be the only other place
 - ch of habitable environments
 - one of the best candidates
- planta
 - En
 - la
 - ierno prevé que
 - ñales la construcción de una
 - 5 percent of French electricity,
- and
 - Wind
- energía
 - biomasa en polideportivos,
 - volucra varios proyectos de
 - La promotora del
 - Uno de ellos es el
- Parque
 -), ya sea en el ámbito
 - **
- 60mm hydrogen-alpha telescope for
- azúcares utilizando la energía de la luz
- deployed, the SEAI failed to mention
- Despite that fact that large scale
- el corto plazo", comentó Alfredo
- Europe is still the world leader in
- experienced with other funds," said
- food, heating with wood, producing
- in France, while electricity obtained from
- ierten Geld in Biogasanlagen, Nahwärmenetze,
- Irish energy firm
- it plans to build a 100 megawatt
- las energías renovables: la española TW
- lo cual, se recomienda utilizar bloqueador
- SANTIAGO
- se produjo durante la infancia del Sistema
- the August 21 st, 2017 total
- the first of five pipeline projects that
- The agreement entails the acquisition of new
- use; underground dripper irrigation systems;
- with over 2.5GW of
- wurde vor ihrer Annexion eine der größten

- 21
 - 's UK-based biomass
 - chief executive Michael Bradley.
 - has agreed with UK-
- energy
 - 1.6 percent,
 - farms that are being develop
 - would triple.
- PV
 - grid connection applications into
 - in the Ireland's Energy Targets
 - use, with over 100
 - beyond Earth.
- system
 - for harbouring life.
 - to have large bodies of
- '
 - con sede en Barcelona y
 - gerente General de SunEdison en
- - Anlagen auf der Welt gebaut
 - oder Windparks und wählten dafür
- developer SunEdison Inc
 - , which
 - SUNEd
-).
- de sa Marina de Lluçmajor desmiente
- eclipse passes almost directly over the
- electricity, improving habitat, writing
- en el salar de Uyuni y
- fotovoltaica en desarrollo, así como
- generará 60 megavatios y estará formada
- is the cheapest form of renewable
- o en el eólico, predominantes
- observing and several Celestron and Meade
- Olmué, ubicado en la comuna
- panels were installed on the roof
- para mezclar el CO2 que
- plant in Chile to supply the
- power get some special tax breaks
- se invertirá unos 94 millones de
- water heating systems; and an
- y zapatos cómodos; y el

abu agreement aircraft airplane astronomers
bankruptcy bathroom battery bill business capital
central china cloud dhabi dining electricity
energies energy flight fuels garage garden
glasses hawaii house ice internet journey kilos
kitchen led manuel mars megawatts mexico
mission moon nasa nature panels plane planets
point pool pot power production pulses question
ray renewable room science service share signing
site smart social space stars sun tower trump

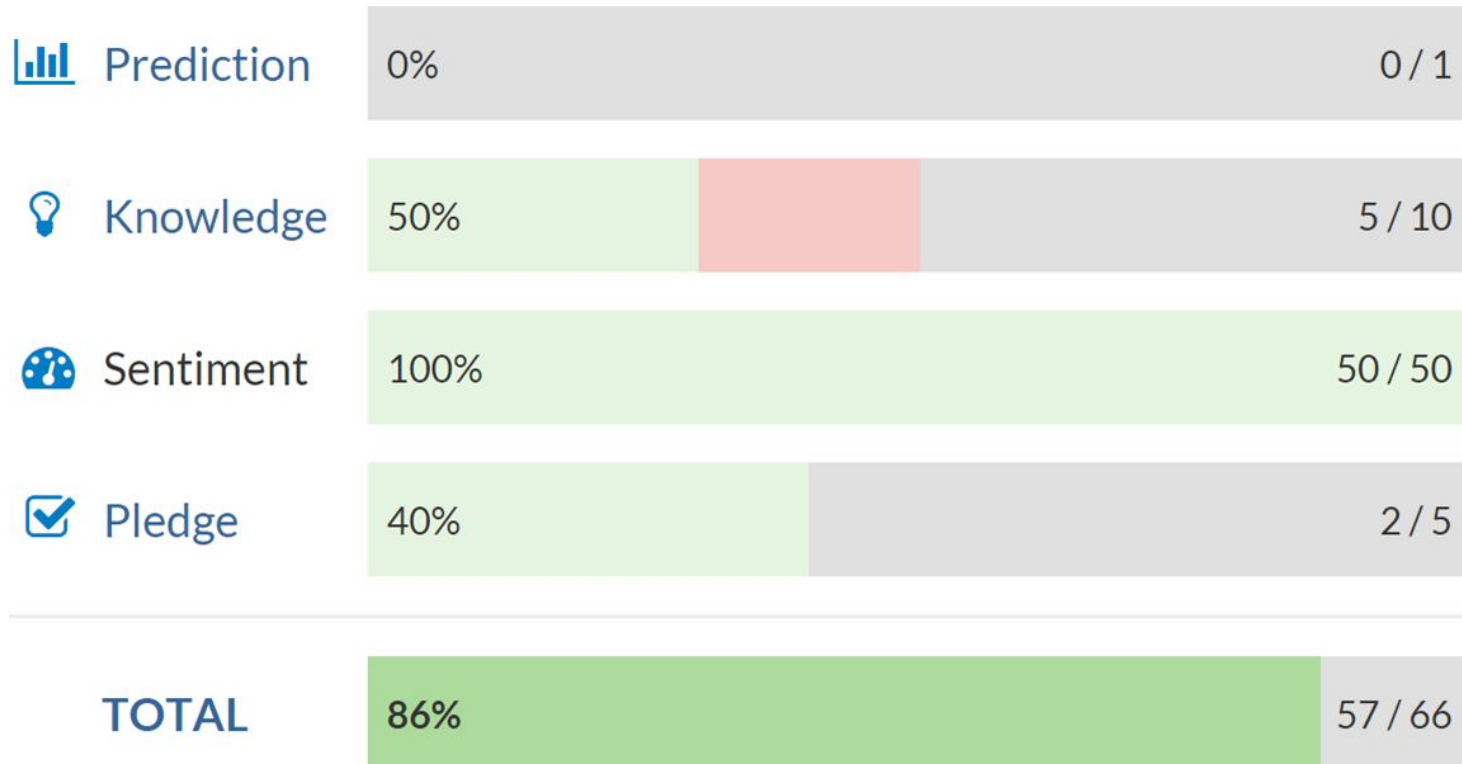


- **Goal and Motivation.** Facilitate GWAP development to engage users and generate valuable information.
 - **Cross-Platform HTML5 Application Framework** including **Social Logins** (Facebook, Twitter, Google+)
 - **Application Programming Interface (API)** that supports hybrid HC approaches (**GWAP** / **CrowdFlower**)
 - **Task Types:** Binary, Multiple Choice, Opinion Polls, Prediction, (Multiple) Sliders
 - **Experiments and Workflow Optimisation**
- **GWAP Applications**

Language Quiz
quiz.ucomp.eu

Climate Challenge
www.ecoresearch.net/climate-challenge

Task Selection



language quiz

Play Progress Leaderboard About

 Sentence Sentiment



Drückt der fett hervorgehobene Text eine **negative**, neutrale oder **positive** Meinung aus?

Zu wenig, sagen die Umweltschutzorganisationen. Gerade noch erträglich, sagen die Bauern. **Denn der Kampf gegen Schädlinge hat sich in Österreich in den vergangenen Jahren verschärft.** Schädlingsplage aus Amerika
Ein fünf Millimeter kleiner schwarz-gelber Käfer sorgt derzeit für Unruhe auf den Äckern: der Westliche Maiswurzelbohrer.



AVAILABLE LANGUAGES

EN	DE	ES	RU	ZH	CS
English	German	Spanish	Russian	Chinese	Czech

Lexicon Acquisition

Climate challenge
test your climate knowledge - predict the future state of our world



[Play](#) [Progress](#) [Leaderboard](#) [About](#)



Sentiment Assessment



Does the following term have a **negative**, neutral or **positive** meaning?

energy conservation



Game Level

1

Level

↑ +5



15



You



105



300



335

Current Questions

OPEN QUESTIONS

- How much warmer or colder than average will sea surface temperature be in April 2015 in the eastern tropical Pacific Ocean (the Niño 3.4 region)?

Top Scores March



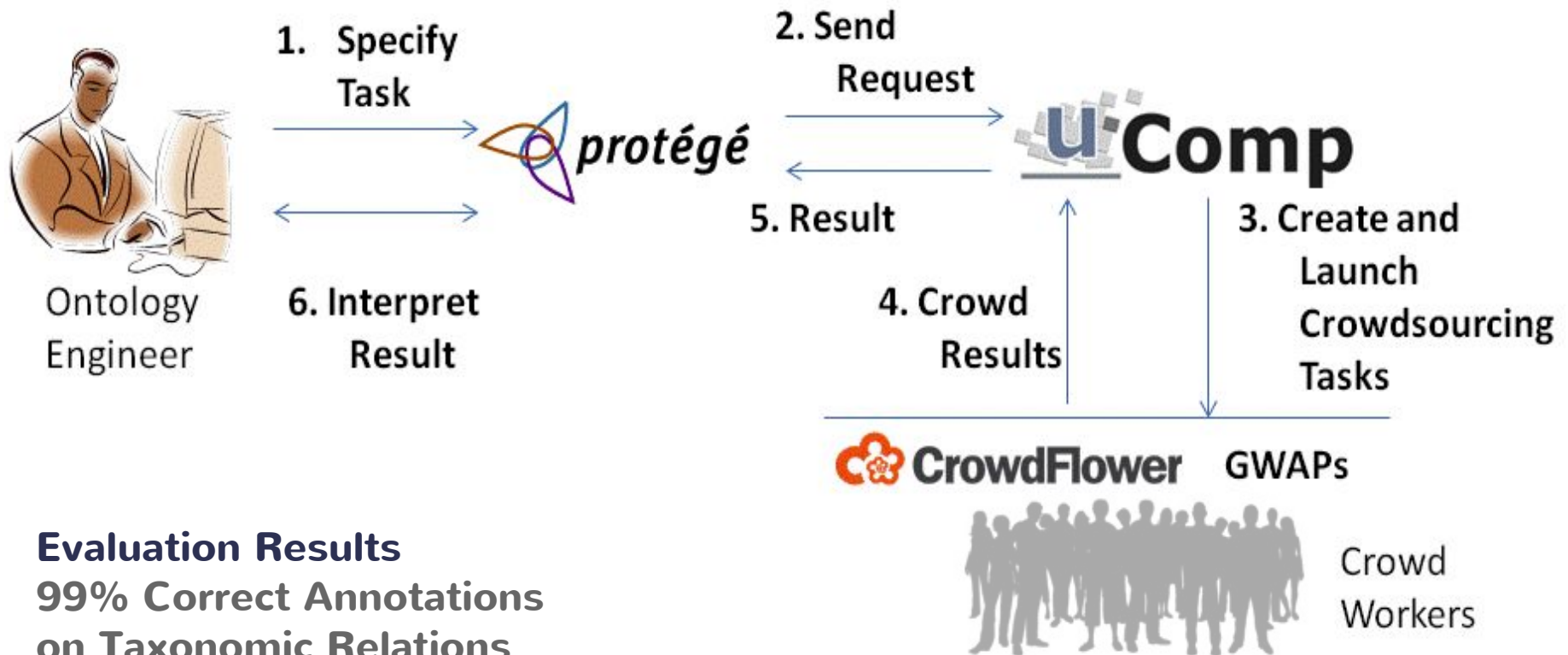
1		Christoph Meili Level: 3	1150
2		Diana Maynard Level: 3	950
8		Antonio Panderas Level: 1	100

- **NLP Task Decomposition** and Mapping to HC Tasks
- **Automated Prioritization** of HC Tasks
 - Automated Relevance Selection
 - Active Learning
- **Recursive Workflows** to Improve Quality
- Hybrid Approaches Combining **Expert Knowledge** and **Collective Intelligence**

- What is the best **Auto-Adjudication Strategy**?
 - “Full Recall” - Take union of crowd annotations
 - 86% average agreement with expert judgements
- Are **Crowd-Annotated Datasets** as good as expert-annotated ones for training ML models?
 - NER experiment using the Stanford NER System
 - 2 Datasets: uComp NE annotated tweets, UMBC
 - NER precision is broadly similar or declines slightly on crowd-annotated data
 - NER recall declines significantly on the crowd-annotated data

- Building on an existing **ontology learning framework** - extended the system to cope with text in multiple **languages** and **domains**
- **New evidence sources**
 - Experiments to select, **balance** and **optimize** the sources
- Integrate **uComp human computation** to verify concept candidates
- Study differences between **domain expert judgements** and **crowd workers**

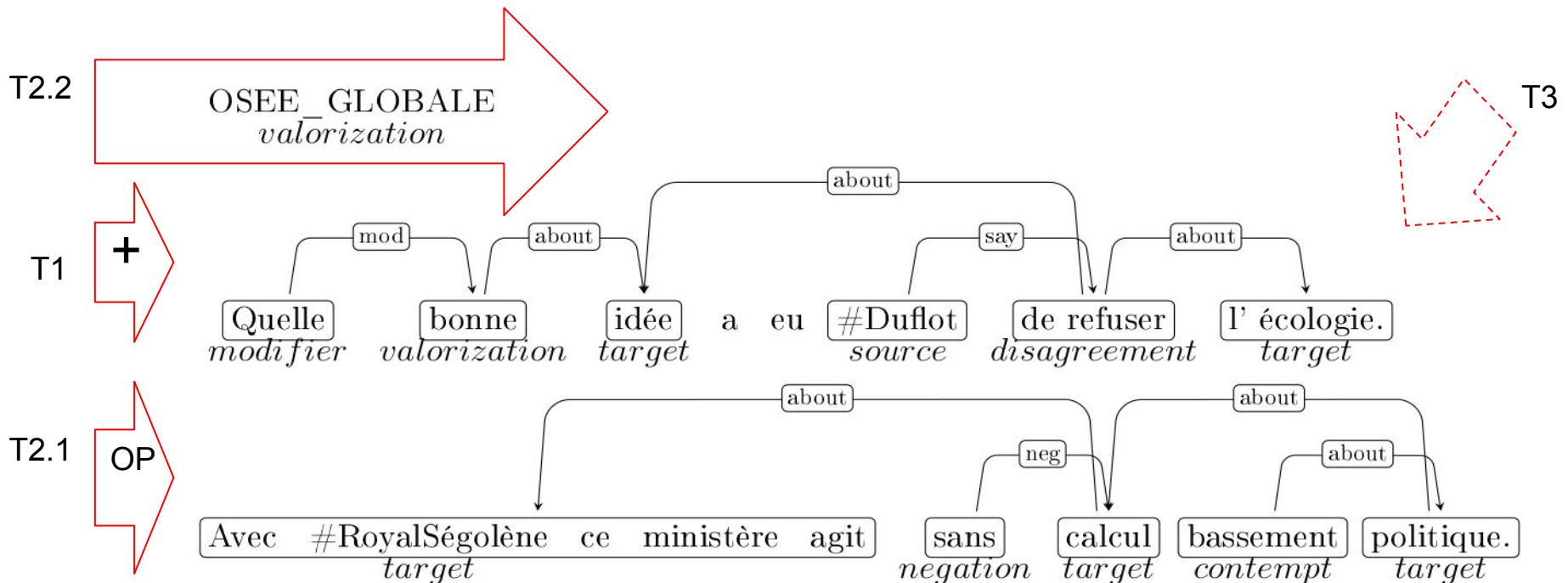
Goal: Apply HC Framework to **Ontology Engineering**



Evaluation Results

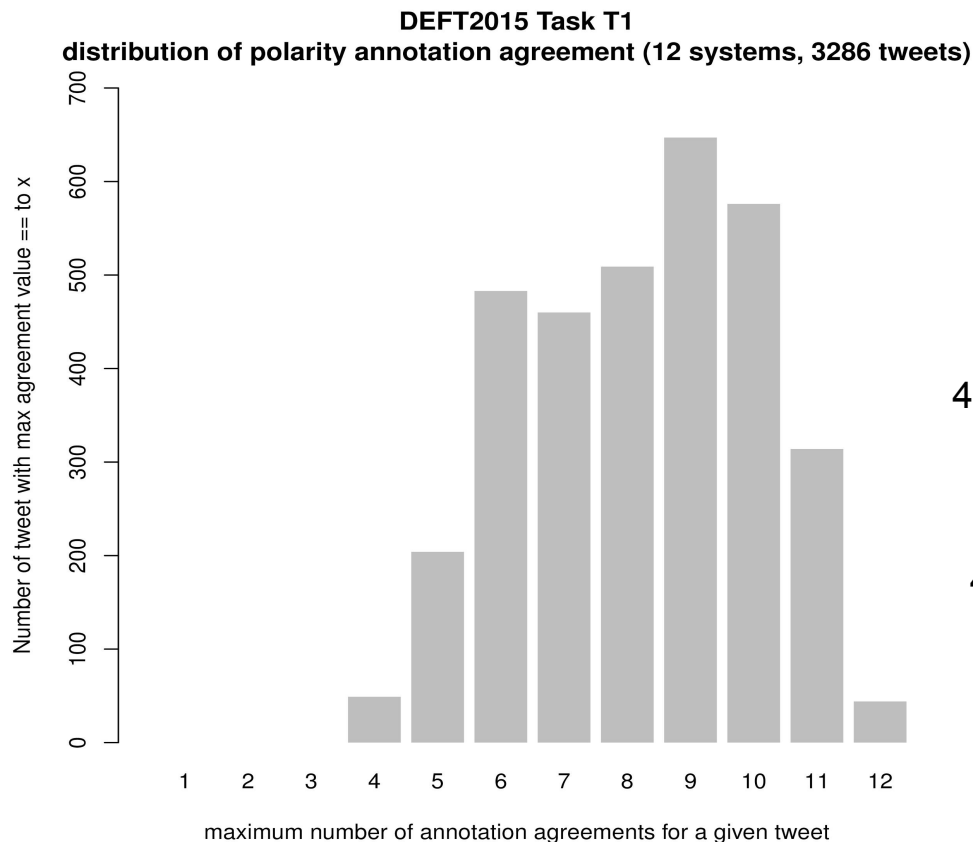
**99% Correct Annotations
on Taxonomic Relations**

- **Crowdsourcing for Shared Task Evaluation**
DEFT 2015 Challenge on French Tweets



ROVER / Active Learning / Sort / Crowdsourcing

Maximum Agreement Value = Most Numerous Cases



4873...593 = + - = + = = = - + = 7

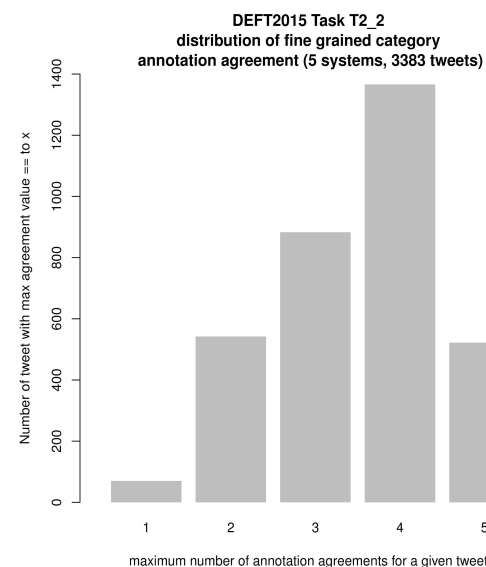
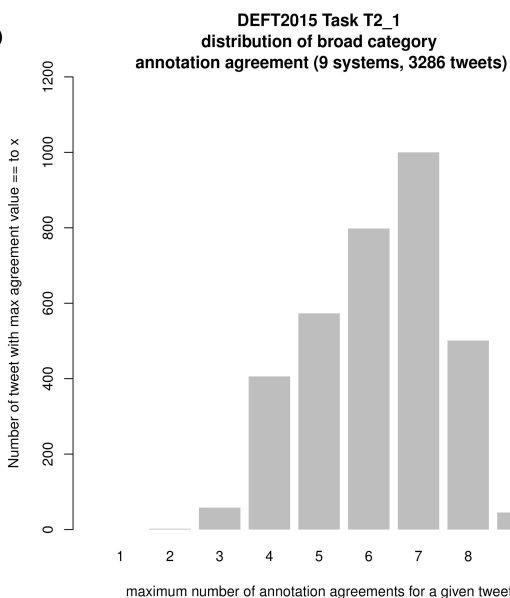
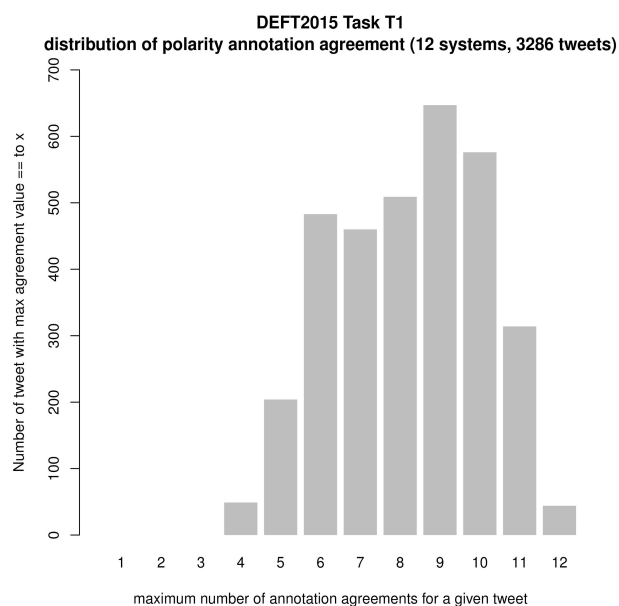
4873...593 INF INF OP INF OP INF INF INF INF 7

4873...593 VALO COLERE VALO VALO INFO 3

- Web Site: www.ucomp.eu; Twitter Presence: [@uCompEU](https://twitter.com/uCompEU)
- Deliverables: 24
Scientific Publications: 38
- Open Source Results
 - Toolkits: eWRT, TwitIE, GATE HC Plugin, Protégé Plugin
 - Datasets: Named Entities (EN); Sentiment (FR, DE); MWCC Content Repository (EN, FR, DE, ES)
- Evaluation Campaign (DEFT 2015)
- Training and Teaching
 - Courses on Mining and Crowdsourcing Social Media Corpora; [GATE Summer School](#) (2014, 2015, 2016)
 - Tutorials at ESWC-2014 and EACL-2014
 - Human Computation exercises as part of the MBA and BBA programs of MODUL University Vienna

ROVER / Active Learning / Sort / Crowdsourcing

Maximum Agreement Value = Most Numerous Cases



4873...593 = + - = + = = = - + = 7

INF INF OP INF OP INF INF INF INF 7

VALO COLERE VALO VALO INFO 3

- **Collaboration :: Scientific**
 - **DecarboNet FP7 | Climate Challenge**
 - **Pheme FP7 | Evaluation**
 - **SoBigData H2020 | Human Computation**
 - **Member of the European Center for Social Media**



- Collaboration :: Societal Impact
 - **United Nations Environment Programme (UNEP)**
GEMET Multilingual Thesaurus, International Visibility

THIS EARTH HOUR, SHINE A LIGHT ON CLIMATE ACTION.

#ChangeClimateChange EARTHHOUR.ORG

19 MAR 2016 | 8:30 PM



- **Collaboration :: Societal Impact**
 - **United Nations Environment Programme (UNEP)**
GEMET Multilingual Thesaurus, International Visibility
 - **WWF Earth Hour**
Climate Challenge Release (03-2015)
Promotion of "Earth Hour Edition" (03-2016)

- **Collaboration :: Societal Impact**
 - **United Nations Environment Programme (UNEP)**
GEMET Multilingual Thesaurus, International Visibility
 - **WWF Earth Hour**
Climate Challenge Release (03-2015)
Promotion of "Earth Hour Edition" (03-2016)
 - **National Oceanic and Atmospheric Administration (NOAA)**
Prediction Task, Climate Resilience Toolkit