

Partners

- **The University of Sheffield**

[GATE team](#), [Department of Computer Science](#)

Regent Court, 211 Portobello, Sheffield S1 4DP, UK

PI [Dr Wim Peters](#) | Co-PI Dr [Kalina Bontcheva](#) | Co-PI [Dr Mark Hepple](#)



- **MODUL University Vienna**

[Department of New Media Technology](#)

Am Kahlenberg 1, 1190 Vienna, Austria

PI [Prof Arno Scharl](#) | Co-PI Dr [Marta Sabou](#)



- **Vienna University of Economics and Business**

[Research Institute for Computational Methods](#)

Augasse 2-6, 1090 Vienna, Austria

PI [Prof Kurt Hornik](#) | Co-PI [Dr Gerhard Wohlgenannt](#)



- **LIMSI-CNRS**

[Man-Machine Communication Dept. ILES group \(Written and Signed Language\)](#)

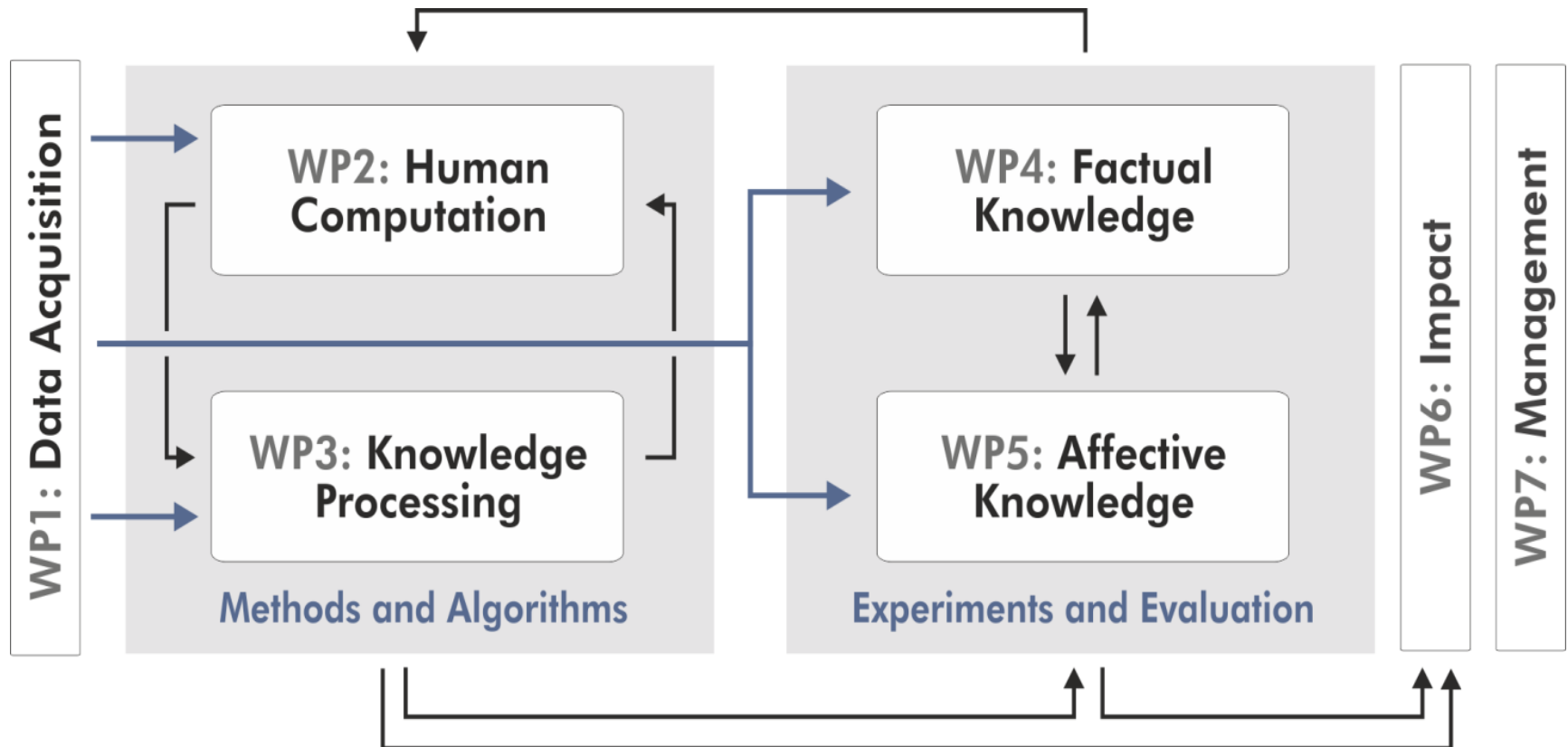
Bat. 508, Univ. Paris-Sud, 91403 Orsay Cedex, France

PI Dr [Patrick Paroubek](#) | Co-PI Dr [Amel Fraisse](#)



- Develop a generic, configurable and reusable **human computation framework**
- Address challenges of **noisy data**
- Embed human computation into **knowledge extraction workflows**
 - Factual Knowledge
 - Affective Knowledge
- **Evaluate** EHC performance
(EHC = Embedded Human Computation)

Work Package Overview



- **Application Framework.** Facilitate developing GWAPs to engage users and generate valuable information.
- **Mechanism.** Players score if **inputs match**: (i) system-generated values; (ii) Real-time input from other players; (iii) stored records from previous users.
- If a certain number of players **agree**, the task will be assumed **complete** and taken out of the game
- **Progress**
 - **HTML5 application framework** to ensure compatibility with mobile platforms. **Complete.**
 - **Application Programming Interface (API) | Complete.**
 - Integration of GWAPs with **CrowdFlower. Ongoing.**

GWAP Use Case



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



[Play](#) [Leaderboard](#) [Rules](#) [FAQ](#)



Sentiment Assessment



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

wind farm



Game Level

1

Level



80



You

Current Questions

ARCHIVE ▾

Top Scores March



1



Arno Scharl

Level: 1

80

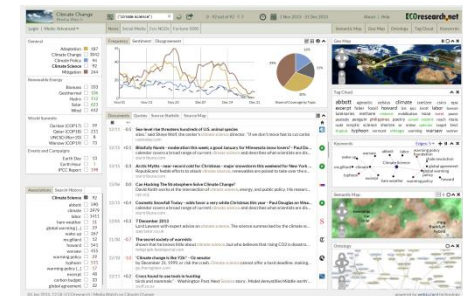
Bonus Points March

0



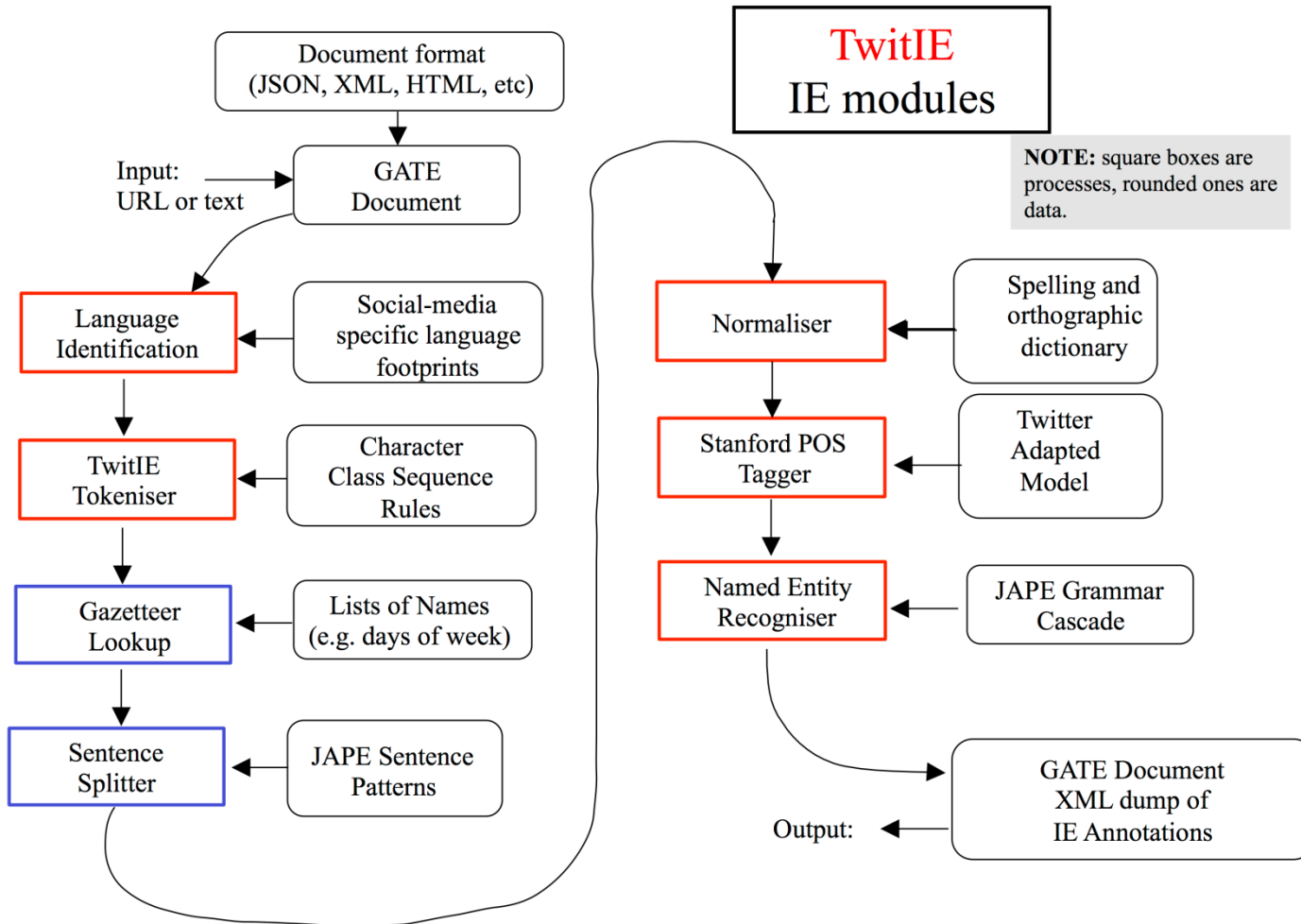
Earn 5% of the points that your friends make after accepting your invitation. The calculation is recursive, so if they invite others you will even get more bonus points.

- Extensible Web Retrieval Toolkit (eWRT)
 - Open Source Library
www.webylard.com/ewrt
- Media Watch on Climate Change
 - English Version
 - www.ecoresearch.net/climate
 - Start: 01 Jan 2013
 - News Media Articles: **215,000**
 - Social Media Postings: **4,110,000**
 - German Version
 - www.ecoresearch.net/climate/de
 - Start: 01 Jan 2013 (News), 01 Sep 2013 (Social)
 - News Media Articles: **142,000**
 - Social Meeting Postings: **123,000**
 - French – Upcoming in April 2014



D1.2 TwitIE (Social Media)

Open-source; download at <http://gate.ac.uk/wiki/twitie.html>



D1.2 TwitIE (Social Media)

System	Precision	Recall	F1
Newswire			
ANNIE	78%	74%	77%
Stanford	-	-	89%
Microblog			
ANNIE	47%	83%	60%
TwitIE	77%	83%	80%
Stanford	59%	32%	41%
Stanford-twitter	54%	45%	49%
Ritter	73%	49%	59%

K. Bontcheva, L. Derczynski, et al. TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text. Proceedings of Int. Conf. on Recent Advances in Natural Language Processing (RANLP). 2013.

TwitIE-as-a-Service

Show only items tagged: [Arabic](#) [Biomed](#) [Dutch](#) [English](#) [French](#) [Genes](#) [German](#) [Hungarian](#) [Hunpos](#) [Measurements](#) [Named Entity](#) [OpenNLP](#) [Parser](#) [Part-of-Speech](#) [Romanian](#) [Stanford Parser](#) [Tagger](#)

ANNIE Named Entity Recognizer  The ANNIE named entity recognition service. Identifies names of persons, locations, organizations, as well as money amounts, time and date expressions. GBP0.09 / MIB data	Arabic Named Entity Recognizer  A named entity recognition service for documents in the Arabic language. Based on ANNIE, it identifies names of persons, locations, organizations, as well as money amounts, time and date expressions. GBP0.09 / MIB data	French Named Entity Recognizer  A named entity recognition service for documents in the French language. Based on ANNIE, it identifies names of persons, locations, organizations, as well as money amounts, time and date expressions. GBP0.09 / MIB data
German Named Entity Recognizer  A named entity recognition service for documents in the German language. Based on ANNIE, it identifies names of persons, locations, organizations, as well as money amounts, time and date expressions. GBP0.09 / MIB data	Romanian Named Entity Recognizer  A named entity recognition service for documents in the Romanian language. Based on ANNIE, it identifies names of persons, locations, organizations, as well as money amounts, time and date expressions. GBP0.09 / MIB data	Measurement expression annotator  Annotates numbers and measurement expressions, with their normalised values in SI units. GBP0.10 / MIB data
ANNIE+Measurements  Annotates named entities (person, location, organization, date) as well as numbers and measurement expressions. GBP0.12 / MIB data	PennBio Tagger  Tokeniser tuned to biomedical text, and taggers for gene, genomic variations and malignancy type entities, from the University of Pennsylvania. GBP0.99 / CPU hour	AbGene Tagger  Tag mentions of gene entities in text using the AbGene tagger. GBP0.99 / CPU hour
OpenNLP English Pipeline  The English tokeniser, sentence splitter, POS tagger, phrase chunker and named-entity recogniser from Apache OpenNLP. GBP0.99 / CPU hour	OpenNLP Dutch Pipeline  The Dutch tokeniser, sentence splitter, POS tagger, phrase chunker and named-entity recogniser from Apache OpenNLP. GBP0.99 / CPU hour	OpenNLP German Pipeline  The German tokeniser, sentence splitter and POS tagger from Apache OpenNLP. GBP0.99 / CPU hour
Hunpos English POS Tagger  Tokeniser, sentence splitter, and part-of-speech tagging for English using the Hunpos HMM-based tagger. GBP0.09 / CPU hour	Hunpos Hungarian POS Tagger  Tokeniser, sentence splitter, and part-of-speech tagging for Hungarian using the Hunpos HMM-based tagger. GBP0.09 / CPU hour	Stanford Parser and POS Tagger  The Stanford Parser, producing POS tags, parse trees and dependency structures for input text. GBP0.09 / CPU hour

K. Bontcheva, L. Derczynski, et al. TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text. Proceedings of Int. Conf. on Recent Advances in Natural Language Processing (RANLP). 2013.

Cloud-based text analytics services on **ANNO-MARKET.COM**

- Open-source, now released as part of GATE
- Download from <http://gate.ac.uk/wiki/crowdsourcing.html>
- Currently two types of tasks:
 - Classification (e.g. entity/word disambiguation, sentiment)
 - Sequence selection (e.g. named entity annotation)
- Tasks commissioned from the GATE Developer UI
- Mapping from sentences/annotations to HC tasks done automatically
- Annotation provenance & contributor reliability tracked
- Collected data mapped back onto corpora and documents automatically

D3.2 GATE HC Plugin

Automatic data pre-processing and mapping to individual tasks

Messages Corpus Pipeline... Blank NE corpus gate_tweet.bc.1...

Loaded Processing resources

Type

Selected Processing resources

!	Name	Type
	Document Reset PR_007E1	Document Reset PR
	ANNIE English Tokeniser_007F3	ANNIE English Tokeniser
	ANNIE Sentence Splitter_007F6	ANNIE Sentence Splitter
	Entity Annotation Job Builder_007DF	Entity Annotation Job Builder

Corpus: Blank NE corpus

Runtime Parameters for the "Entity Annotation Job Builder_007DF" Entity Annotation Job Builder:

Name	Type	Required	Value
entityASName	String		
entityAnnotationType	String	✓	Person
goldFeatureName	String	✓	gold
goldFeatureValue	String	✓	yes
jobId	Long	✓	
snippetASName	String		
snippetAnnotationType	String	✓	Sentence
tokenASName	String		
tokenAnnotationType	String	✓	Token

Auto-Created Sequence Selection



Click to mark the words that are part of location names

In each sentence below, mark any names that are locations (e.g. **France**). Don't mark locations that don't have their own special name.

There may be no locations in the sentence at all - that's OK.

Come on folks of # wigan True r False there 's a nutter hanging about wigan with a gun. Darlington st area ?

After marking: (required)

- ☒ All the location names in this sentence are now marked
- ☐ This sentence contains no proper location names

Complete **Tosca** on the tube <http://t.co/O90deSLB>

Arizona collector car auctions bring in \$223.8 million up 21.7% from 2012
<http://t.co/rfHYvT3e> retail therapy run amok?

Previous boundary Next boundary ☐ Overl

Context e Tosca on the tube <http://t.co/O90deSLB>

Mention 

Features

cf_judgment=1128439529, trust=0.904411764705882, worker_id=218

options

{<http://dbpedia.org/r>
drama by the 19th-century french playwright victorien
Sardou. It was first performed on 24 November 1887 at the
Théâtre de la Porte Saint-Martin in Paris, with Sarah
Bernhardt in the title role. Despite negative reviews from
the Paris critics at the opening night, it became one of
Sardou's most successful plays and was toured by Bernhardt
throughout the world in the years following its premiere.
The play itself is no longer performed, but its operatic
adaptation, Giacomo Puccini's Tosca, has achieved enduring
popularity. There have been several other adaptations of the
play including two for the Japanese theatre and an English
burlesque, Tra-La-La Tosca (all of which premiered in the
1890s) as well as several film versions. La Tosca is set in
Rome on 17 June 1800 following the French victory in the
Battle of Marengo. The action takes place over an eighteen-
hour period, ending at dawn on 18 June 1800. Its melodramatic
plot centers on Floria Tosca, a celebrated opera singer;
her lover, Mario Cavaradossi, an artist and Bonapartist
sympathiser; and Baron Scarpia, Rome's ruthless Regent of

cf_unit 372655845

- Use HC to produce **affective resources** that are difficult to obtain automatically and too costly to produce manually, for multiple languages (**EN, FR, DE**).
- Assess HC-produced resources by evaluating the performance impact of using them instead of traditional resources for opinion mining and sentiment analysis (**quantitative black-box methodology**).
- Assess the possibility to replace static **gold standard resources** by dynamic HC

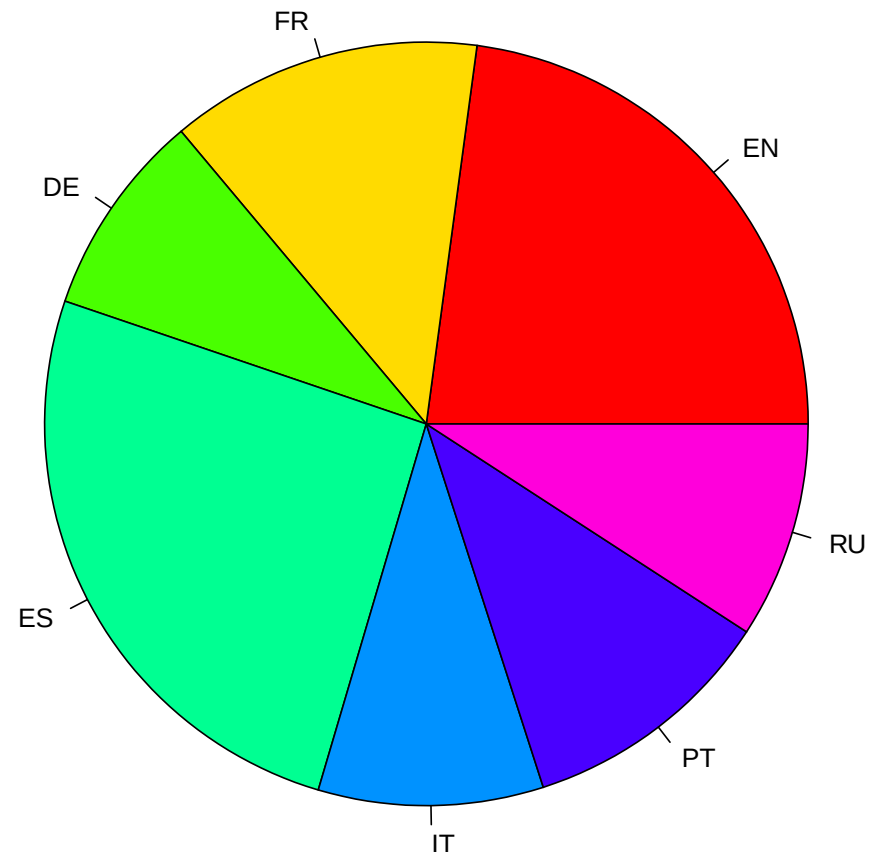
- **The OSE Model - a global and generic model with 3 subjective levels:**
 - Intellective states: Opinions
 - Intellective-affective states: Sentiments
 - Affective states: Emotions
- **Data Acquisition**
 - Step 1: Use Social Media to acquire affective corpus
 - Step 2: Automatic Extraction of affective seed lexicons.
 - Step 3: Use UC framework to validate and extend incrementally the affective lexicons.

#	Label	Dim.	uComp Semantic Category
1	NEGATIVE SURPRISE	e-	negative surprise / negative amazement
2	DISCOMFORT	e-	discomfort / disturbance / embarrassment / guilt
3	FEAR	e-	shyness / worry / apprehension / alarm fear / terror
4	BOREDOM	e-	boredom
5	DISPLEASURE	e-	displeasure / deception / abuse
6	SADNESS	e-	sadness / resignation / despair / sorrow / hopelessness
7	ANGER	e-	impatience / annoyance / irritation / nervousness / anger / exasperation
8	CONTEMPT	e-	reluctance / contempt / disdain / blame / disgust / hate
9	DISSATISFACTION	s-	disappointment / dissatisfaction's / discontent / shame
10	DEVALORIZATION	o-	disinterest / devalorization / depreciation
11	DISAGREEMENT	o-	disapproval / disagreement
12	VALORIZATION	o+	interest / valorization / appreciation
13	AGREEMENT	o+	understanding / approval / agreement
14	SATISFACTION	s+	satisfaction / contentment / pride
15	POSITIVE SURPRISE	e+	positive surprise / positive amazement
16	APPEASEMENT	e+	relief / appeasement / peacefulness forgiveness / thankfulness
17	PLEASURE	e+	pleasure / entertainment / enjoyment / joy / happiness / euphoria / play
18	LOVE	e+	love / affection / care / tenderness / fondness / kindness / attachment / devotion / passion / envy / desire
19	INFORMATION	i	information / announcement / news / demand / query / question
20	INSTRUCTION	i	recommendation / suggestion / instruction / order /command

Affective Corpus

	#Ukraine #death toll rises as clashes continue #sad #grief.
	#Ukraine 60 morts aujourd'hui !!! c'est vraiment #triste #chagrin
	Stop the #violence in #Ukraine 60 #tod heute #traurig
	Impresionantes imagenes de #Kiev que pasaran por fotogramas de una pelcula de guerra !! mu- chos #muertos !! estoy #triste .

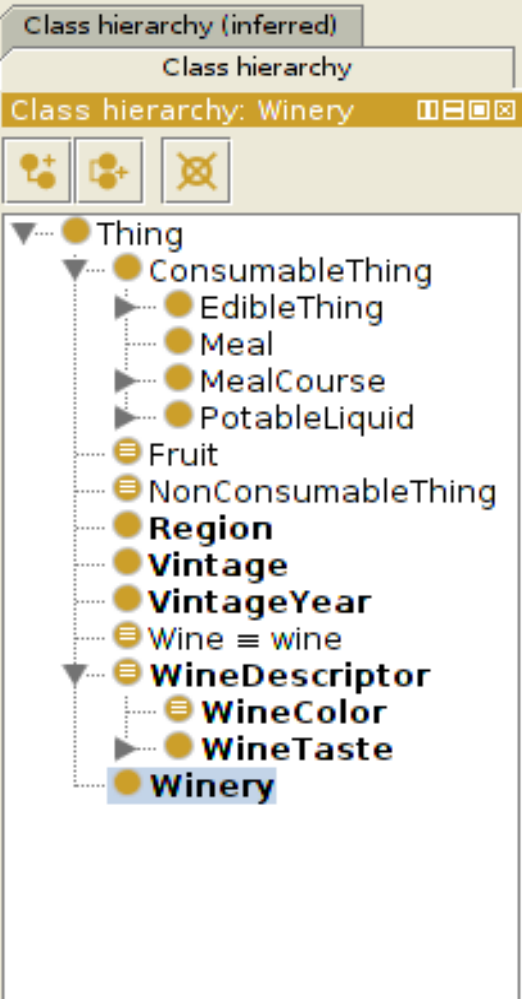
Affective Corpus (English hashtag translation) 59193 tweets



- Ontologies create **shared meaning** and are a cornerstone of the **Semantic Web**
- **Manual construction** of ontologies is cumbersome and expensive
- **Ontology learning** is a (semi-)automatic process to assist the ontology engineer
- uComp builds on an existing ontology learning **framework**

- **Protégé** is the a popular ontology engineering platform
- **Goal:** apply our HC framework in ontology learning and other ontology construction tasks
- **How:** a plugin implemented for Protégé which uses the uComp HC API in order to validate ontological entities

- **Concepts:**
Is the concept relevant for the domain?
- **SubClassOf relations:**
Is concept X a subClass of concept Y?
- **InstanceOf relations:**
Is X an instance of Y?
- **Domain and Range validation:**
Does property Z have a subject X or/and an object Y?
- Suggest **labels** for unlabeled relations
(for automatically learnt ontologies)



Annotations Usage GWAP Class Validation

GWAP Class Validation: Winery

Concept to be validated: Winery

Validate relevance for domain: Wine

Additional information for validators: Is this concept relevant for the domain?

☒ Validate subtree GO ☐ Cancel subtree CANCEL

Description: Winery

Equivalent To +

SubClass Of +

SubClass Of (Anonymous Ancestor)

Members +

Bancroft	?	@	x
Beringer	?	@	x
ChateauChevalBlanc	?	@	x

wine (http://www.w3.org/TR/2003/CR-owl-guide-20030818/wine)

Search for entity

Annotation Properties

Individuals

OWLViz

DL Query

OntoGraf

SPARQL Query

Ontology Differences

Active Ontology

Entities

Classes

Object Properties

Data Properties

Object property hierarchy:   



- topObjectProperty
 - adjacentRegion
 - course
 - hasDrink
 - hasFood
 - hasMaker
 - hasVintageYear
 - hasWineDescriptor
 - locatedIn
 - madeFromFruit
 - madeIntoWine
 - producesWine

Annotations

Usage

GWAP Object Property Validation

GWAP Object Property Validation: locatedIn

In the domain of: Wine

Object Property to be validated: locatedIn

Validate relevance for Subject (Domain): Thing

Validate relevance for Object (Range): Region

Additional information for validators: Please validate domain and range.

GO

CANCEL

Waiting for results from GWAP...please check again later...

Character  

- ☐ Functional
- ☐ Inverse func
- ☒ Transitive
- ☐ Symmetric

Description: locatedIn

Domains (intersection) +

Thing

Ranges (intersection) +

Region

No Reasoner set. Select a reasoner from the Reasoner menu ☒ Show Inferences

- Project Web Site: www.ucomp.eu
- Twitter Presence: [@uCompEU](https://twitter.com/uCompEU)
- Deliverables (8): D1.1, D1.1.1, D1.2, D1.2.1, D2.1, D3.1, D3.2, D5.1
- Open-Source Toolkits (3): eWRT, TwitIE, Gate HC Plugin
- Publications: Scientific Articles (16); Media Coverage (10)
- Collaboration: DecarboNet (Climate Challenge), PHEME (Evaluation), Member of the European Center for Social Media
- Training and Teaching
 - Tutorial: NLP for Social Media. 14th Conference of the European Chapter of the Association for Computational Linguistics ([EACL-2014](#))
 - Week-long course on Mining and Crowdsourcing Social Media Corpora. Annual [GATE Summer School](#) (9 - 13 June 2014)
 - The 6th GATE Training Course (3-7 June 2013, Sheffield, UK). Module on mining social media, based on TwitIE.