

# Crowd-based Ontology Engineering with the uComp Protégé plugin

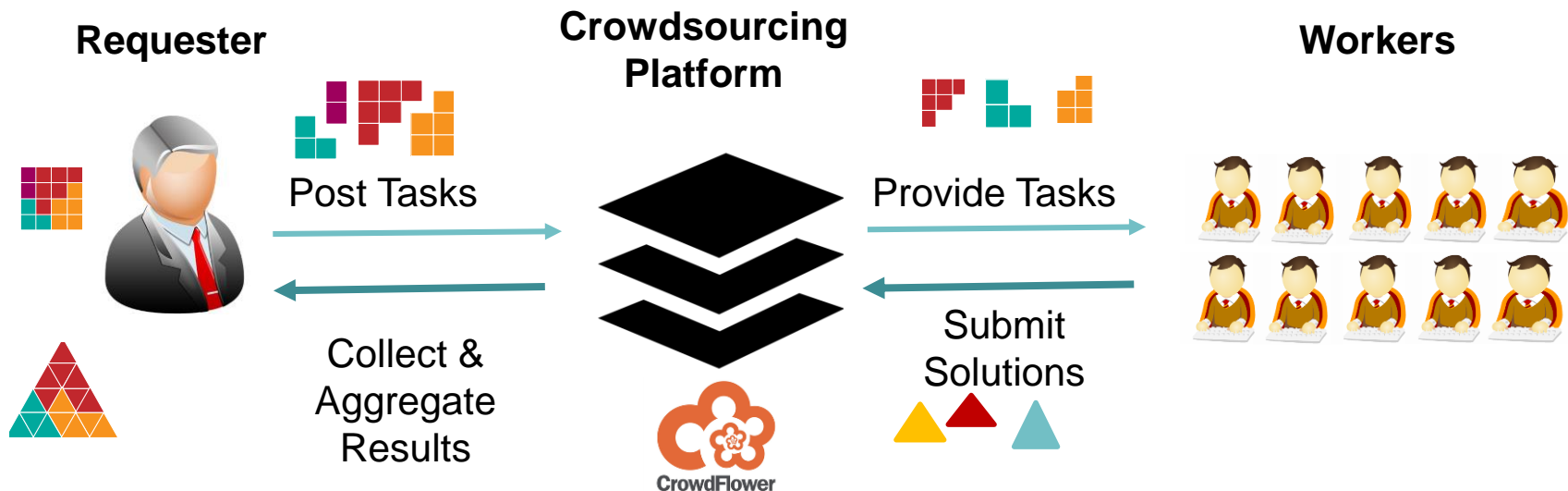
*Gerhard Wohlgenannt<sup>1</sup>, Marta Sabou<sup>2</sup>, Florian Hanika<sup>1</sup>*

*Vienna University of Economics and Business (WU)<sup>1</sup>,  
Technical University of Vienna (TU Wien)<sup>2</sup>*



*G. Wohlgenannt, M. Sabou, F. Hanika: Crowd-based ontology engineering  
with the uComp Protégé plugin. Semantic Web 7(4): 379-398 (2016)*

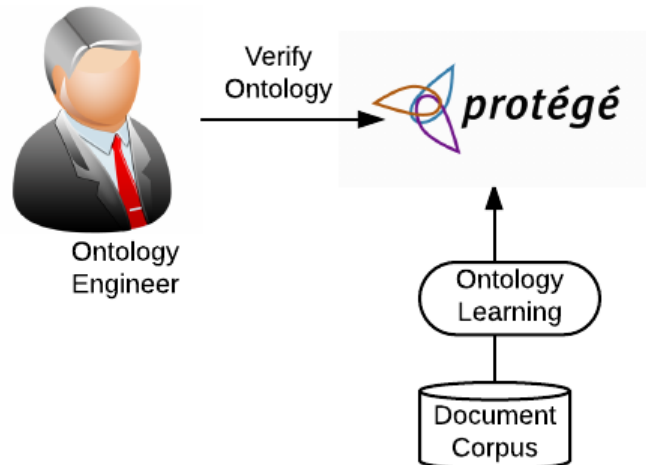
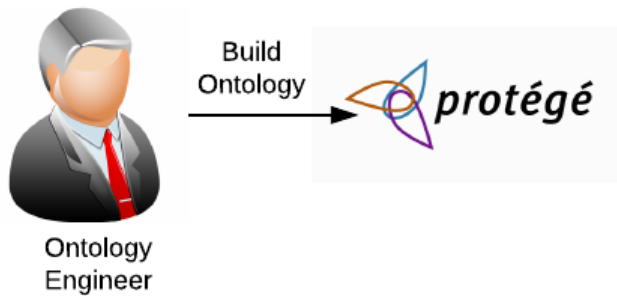
“Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an **undefined**, generally **large** group of people in the form of an **open call**. (J. Howe)”



# Crowdsourcing is a promising approach for addressing ontology engineering problems


Ontology engineering is time consuming and costly.



Crowdsourcing was used successfully in various ontology engineering settings.




Approach	Genre
Verbosity [33]	GWAP
Common Consensus [17]	GWAP
Categorilla [31]	GWAP
Free Association [31]	GWAP
InPho [9]	MLab
Noy [19]	MLab
OntoPronto [27]	GWAP
Climate Quiz [26]	GWAP
Guess What?! [18]	GWAP

facebook Search Arno Scharl Home




Play Leaderboard Community About



Switch Concepts

environmental activism

is a sub-category of

activism


Submit Answer









Skip this Question

**Game Notifications**

Correct. 100% of players agreed


Points: ▲ +1


**Top Scores April 2012**

1		<b>Ernest Rudel</b> Level: 8	100
2		<b>Matt Dearborn</b> Level: 6	61
3		<b>Tigre Timide</b> Level: 5	42
4		<b>Thomas Brudertom Jr.</b> Level: 4	21
5		<b>Arno Scharl</b> Level: 4	18
6		<b>Jan Peters-Anders</b> Level: 1	2
7		<b>Florian Winter</b> Level: 1	2

**Level Status**

4  
Level





18

21

▲  
You

The „Climate Quiz“ is a Facebook application in the tradition of Games with a Purpose. By playing the game, you contribute to a scientific project that aims to raise awareness and share environmental knowledge. You will face two types of challenges: (i) select correct relations between two concepts; (ii) answer climate-change related questions. [<more>](#)

**Friends Invited: 3**

**April 2012 Bonus Points: 0**

▶ **Tell your Friends!**

You will earn **10% of the points a friend makes** after accepting your invitation! The calculation is recursive, so if this friend invites another one you will even get more bonus points.

Ontology engineering is time consuming and costly.

Crowdsourcing used successfully in various ontology engineering settings.

Crowdsourcing requires high upfront investment and therefore remains outside the reach of most ontology engineers.

## OUR MISSION

How to embed crowdsourcing into ontology engineering?

Which ontology engineering tasks can be crowdsourced?

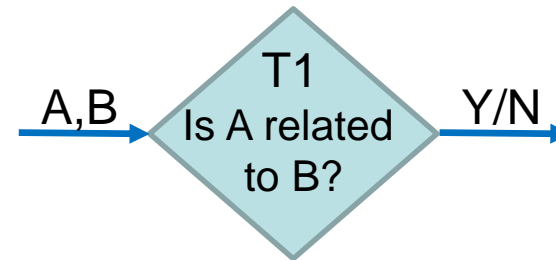
How to implement crowdsourcing enabled ontology engineering ?

Is crowd-based ontology engineering feasible and scalable?

# Ontology engineering problems can be broken down into a set of recurring tasks

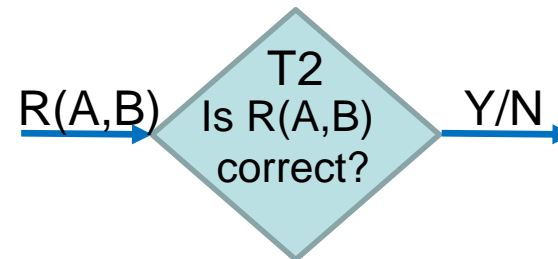
## T1. Specify Term Relatedness

FreeAssociation, InPho, SpotTheLink



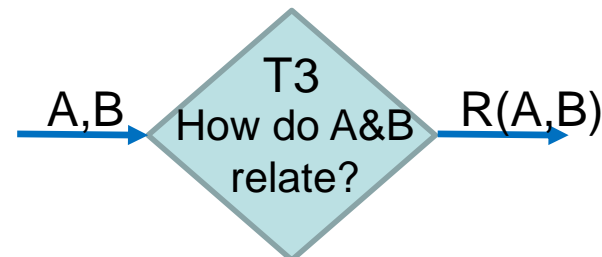
## T2. Verify Relation Correctness

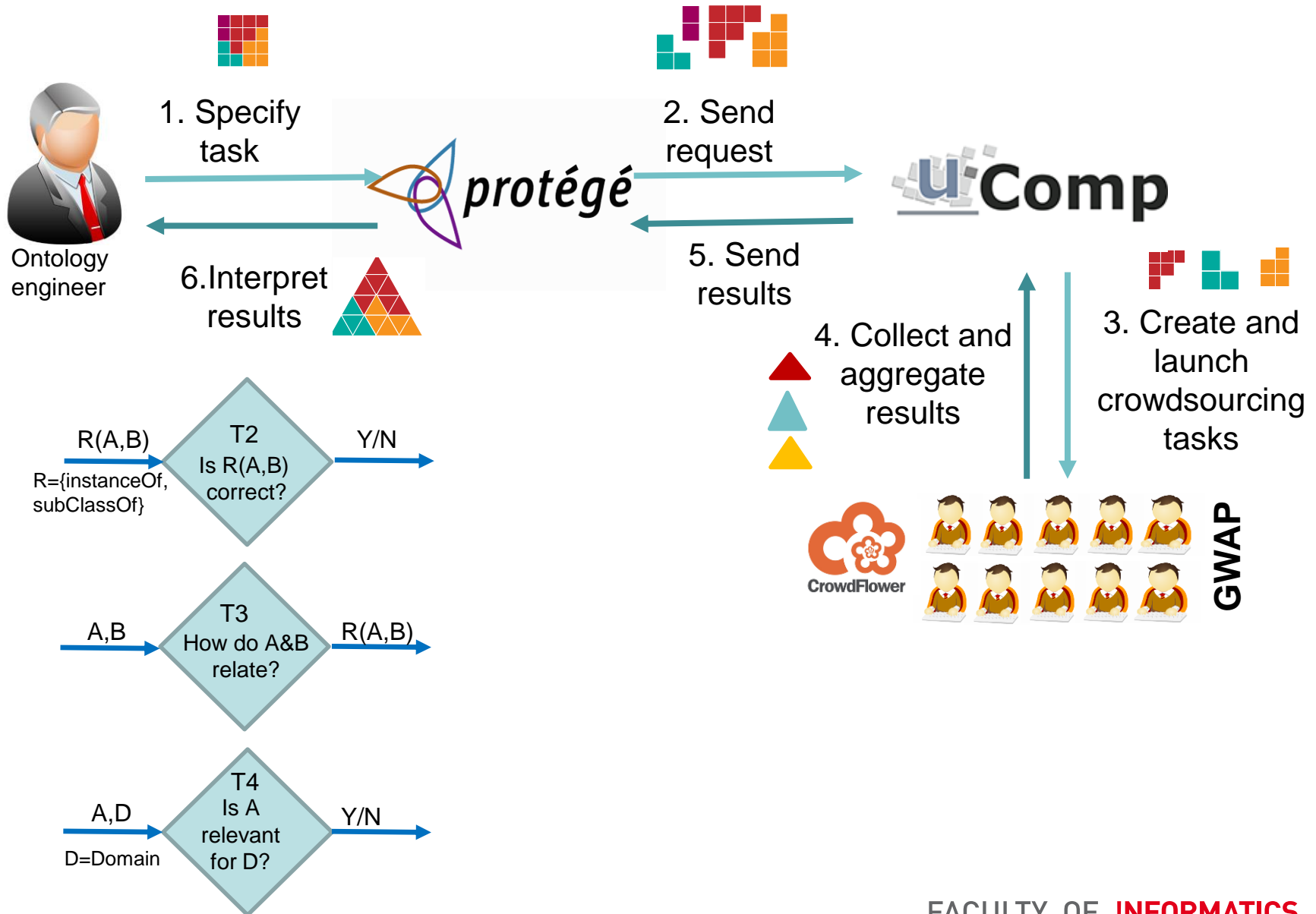
Noy[19], CrowdMap, Chetham[5]



## T3. Specify Relation Type

Verbosity, Categorilla, InPho, OntoPronto, ClimateQuiz, CrowdMap, SpotTheLink





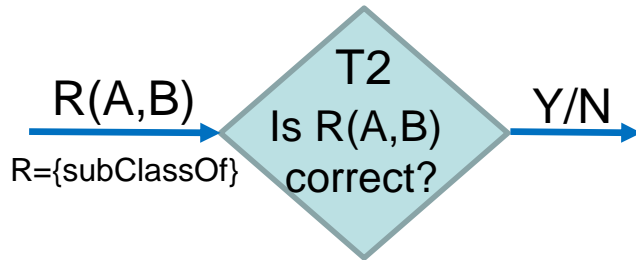


## Verify That A Term Is More Specific Than Another

Instructions ▾

In the domain of Climate Change: Is class electricity a subclass of energy?  
(required)

- Yes  
 No






Data & Tasks

	Ontologies	Classes	ReIs.	Inst.
Feasibility	Climate Change	101	77	0
	Finance	77	50	0
	Wine	138	228	206
	Tennis	52	67	0
Scalability	Human	3304	3761	0

Tasks	Explanation
T4	Judge domain relevance
T2	Verify subsumption correctness
T2	Verify instnaceOf correctness
T3	Specify relation Type


Population

Manual OE




8 Ontology engineers

OE with the Protégé plugin



8 Ontology engineers

Level 3 CrowdFlower workers from English speaking countries 

5 units/task, \$0.05/task, 5 judgements/unit

Metrics

Time  $T_{tt} = T_{oe}$

Cost  $C_{oe} = T_{oe} * 52K\$/Y$

Quality

$T_{tt} = T_{oe} + T_{crowd}$

$C_t = C_{oe} + C_{crowd}$

	Feasibility	Scalability
$T_{oe}$	avg. 90% OE time reduction	-
$T_{tt}$	10 fold increase (21min=>217min)	Similar*
$C_{oe}$	avg. 89% OE cost reduction	-
$C_t$	40% cost reduction (up to 83% with more economic CF settings)	75% - 80%
Quality	comparable with expert judgements	Accuracy: 89%-99%

	Domain Verification		Subsumption Verification	
	Crowd	Est. Manual	Crowd	Est. Manual
Time (H/Day)	19/0.8	19.65/2.5	136/5.6	39.20/4.9
Cost (\$)	104+26TF	511	155+39TF	1019
Quality (Accuracy)	0.99	-	0.895	-

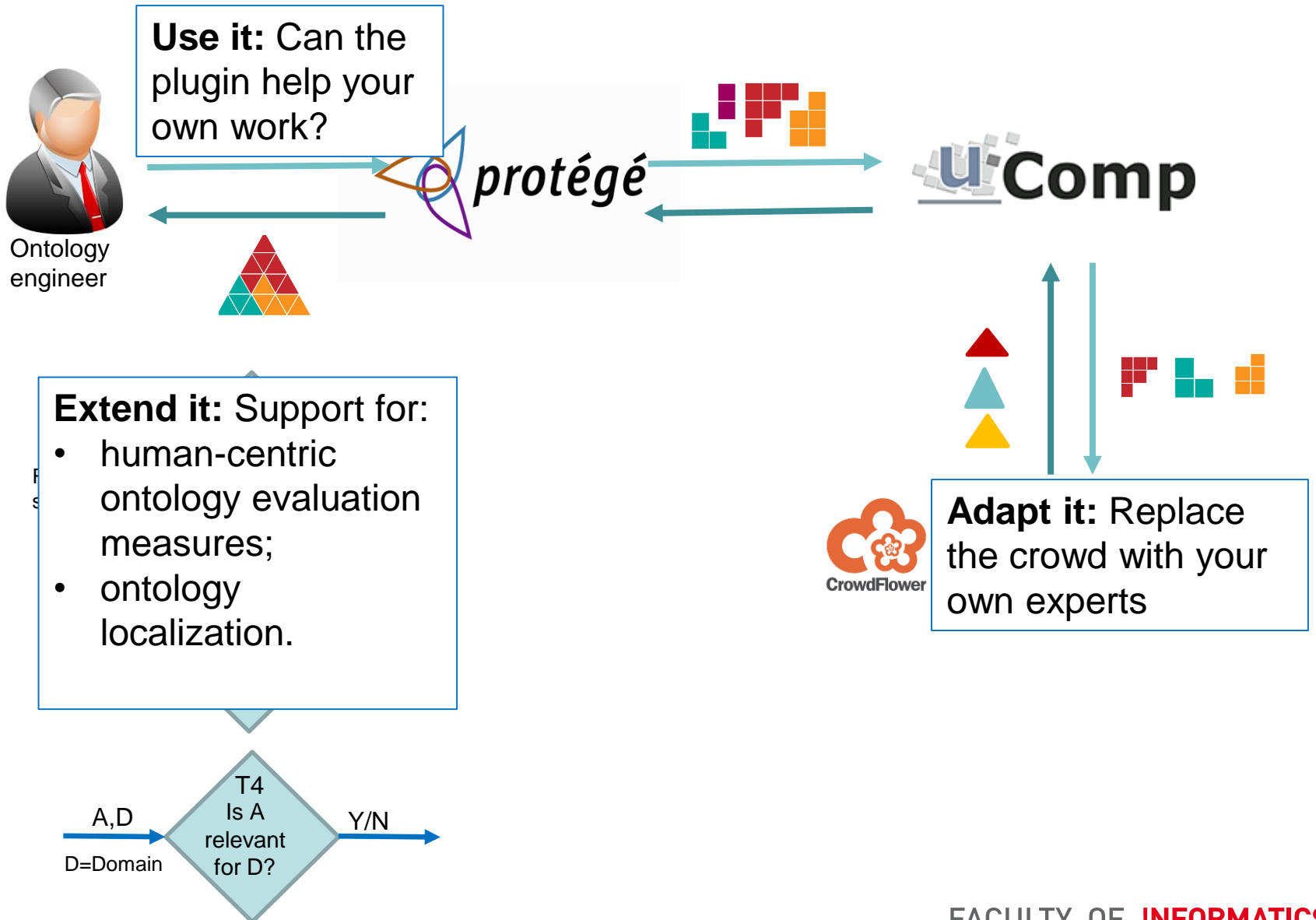
Crowdsourcing requires high upfront investment and therefore remains outside the reach of most ontology engineers.

Crowdsourcing can be embedded successfully in ontology engineering settings.

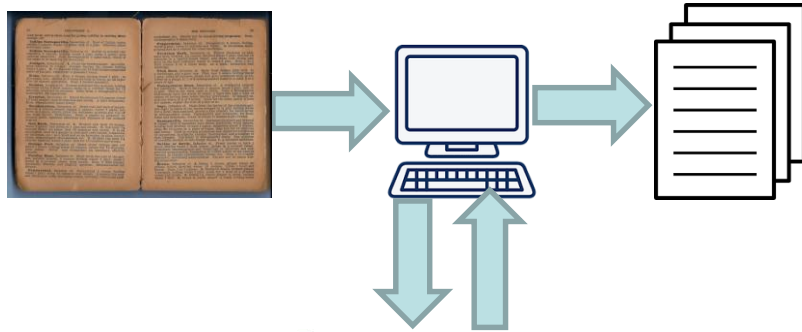
Ontology engineering problems can be broken down into a set of recurring tasks.

uComp Protégé plugin is a prototype for crowd-based ontology engineering.

*Good quality* ontologies obtained *cheaper* and by reducing the time involvement of ontology engineers.



# Human Computation harnesses Human Contributions on Large Scale



100 million/  
day

Digitization of:  
13 million NYT articles;  
2 million books per year;



“Players took 3 weeks to solve the three dimensional structure of a simian retroviral protein that is used in animal models of HIV, but whose structure had eluded biochemists for more than a decade.”

<http://blogs.nature.com/spoonful/2012/04/foldit-games-next-play-crowdsourcing-better-drug-design.html>

S. Cooper, [other authors], and **Foldit players**: Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756-760, 2010.