



## WP7: Patents Use Case

Merítxell Gonzàlez Bermúdez

3rd Project Meeting  
Helsinki, 31st August

# Objectives

- To create a prototype of MT and retrieval of patents in the bio-medical & pharmaceutical domains.
  - Allowing translation of patent abstracts & claims in at least 3 languages.
  - Exposing several cross-language retrieval paradigms on top of them.

# Tasks & Planning

# PM

Partners	PM	Tasks
UPC	15	<ul style="list-style-type: none"><li>• Corpus building</li><li>• Hybrid translation</li><li>• Evaluation</li></ul>
Ontotext	15	<ul style="list-style-type: none"><li>• Semantic infrastructure</li><li>• Prototype building</li></ul>
UGOT	12	<ul style="list-style-type: none"><li>• Domain grammar</li></ul>

# Deliverables

No.	Title	Date
D7.1 Prototype	Patent MT and Retrieval Prototype Beta	M21
D7.2 Prototype	Patent MT and Retrieval Prototype	M27
D7.3 Report	Patent Case Study Final Report	M33

# Tasks

TASK	Name	Leader	Participants
7.1	User Requirements	All	WP9
7.2	Corpora	UPC	WP5
7.3	Grammars	UGOT	
7.4	Ontology	Ontotext	UGOT
7.5	Patents Retrieval System	Ontotext	
7.6	Machine Translation System	UPC	UGOT
7.7	Prototype (user Interface)	Ontotext	UGOT & UPC (MT)
7.8	Evaluation	EHU	All
7.9	Reports	UPC	All

# Calendar

May (M15)	June (M16)	July (M17)	August (M18)
	T7.1. User Req (WP9) T7.2. Corpora - MAREC Corpus	T7.7. Beta Prototype – Architecture	T7.2 Corpora - Compounds list and patents translated in the three languages
	T7.3. Grammar - preliminary report	T7.4 Ontology – Queries form	T7.3 Grammar - Queries grammar
September (M19)	October (M20)	November (M21)	December (M22)
D5.1. Corpora T7.3. Grammar - final report			
T7.7. Prototype – Beta Interfaces and integration		D7.1. Beta Prototype	
January (M23)	February (M24)	March (M25)	April (M26)
		D5.2. Hybrid prototype	

MOLTO

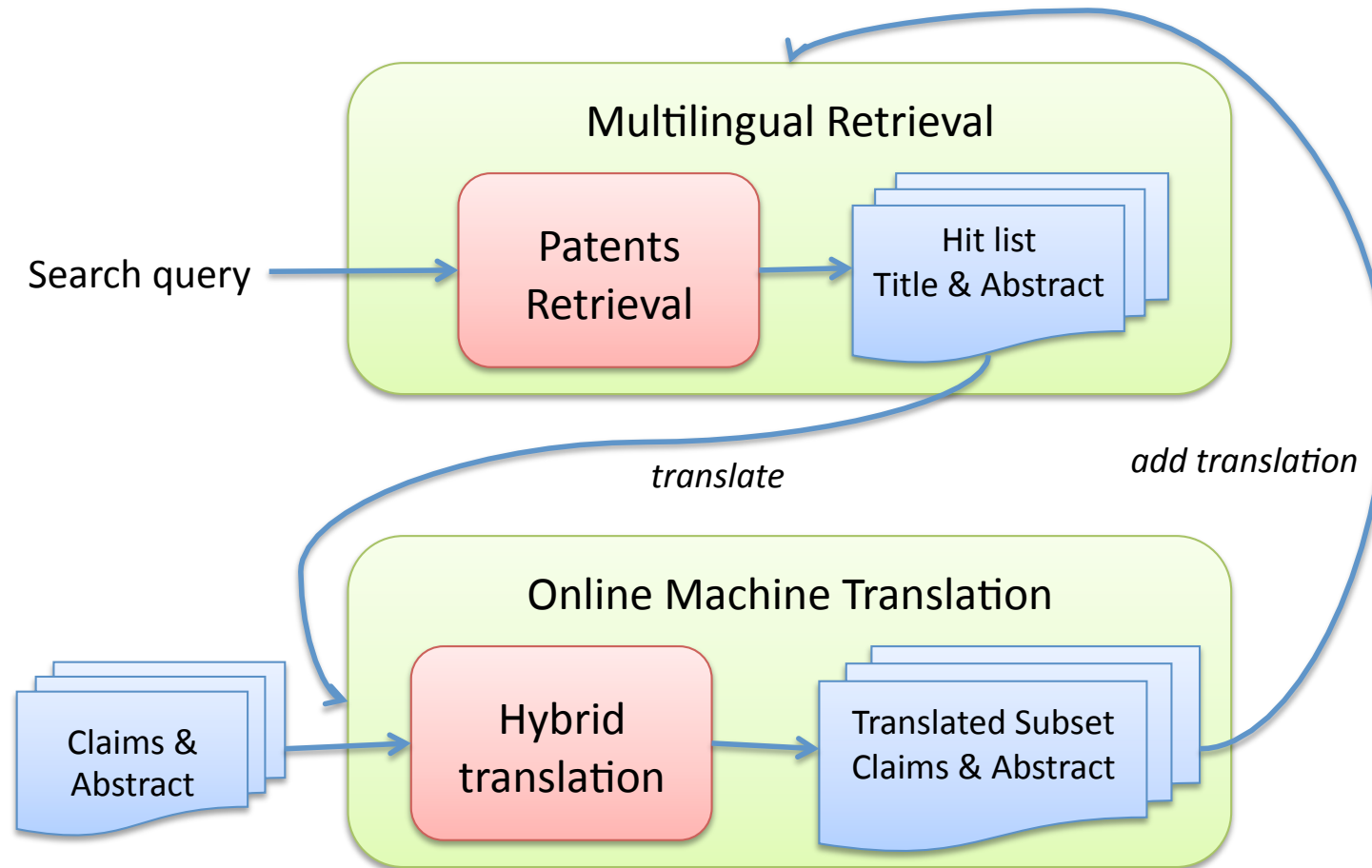
Ongoing work



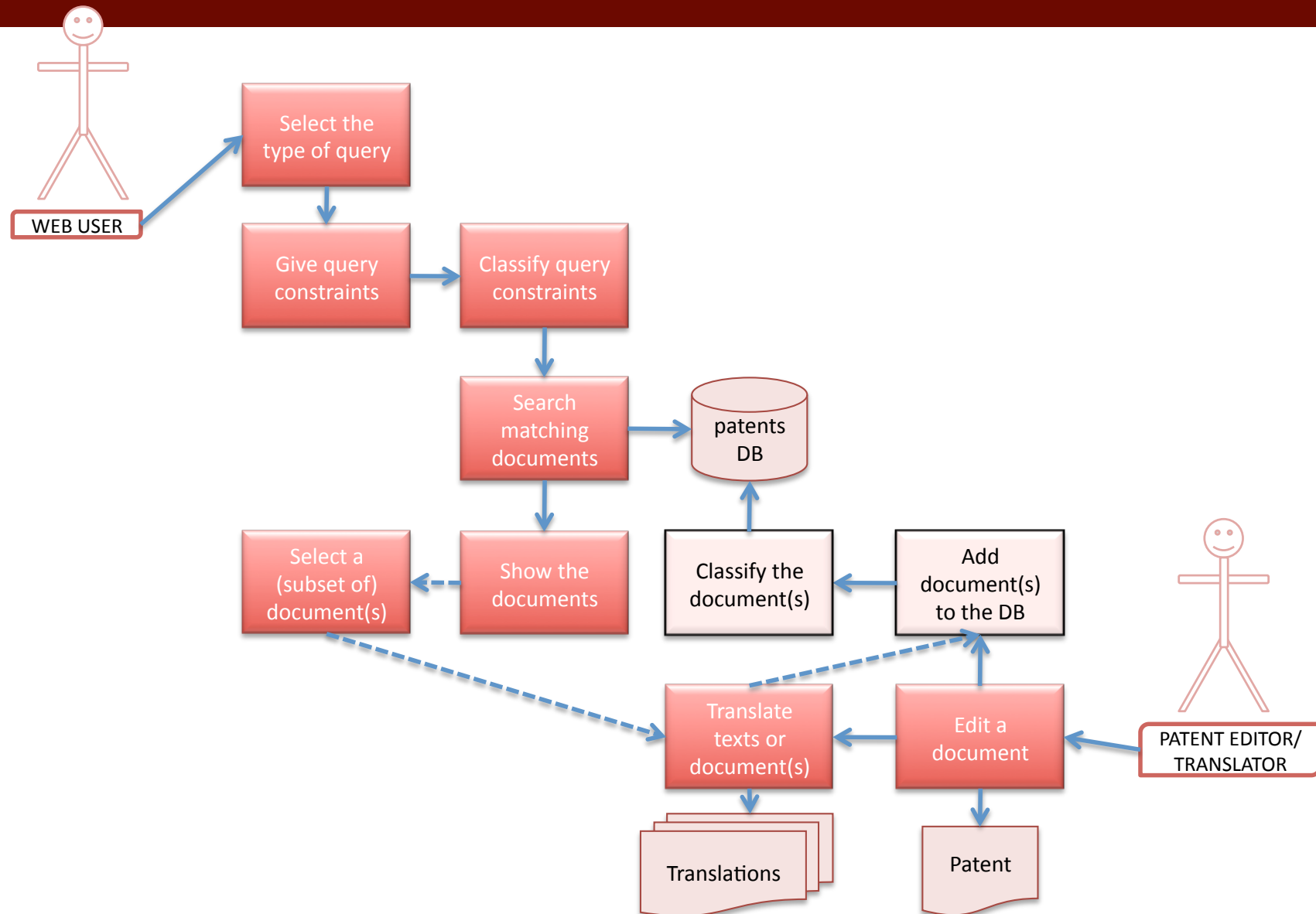
# Overview

Partner	Tasks
UPC	<ul style="list-style-type: none"><li>• Architecture definition</li><li>• User requirements and scenarios</li><li>• Corpus analysis and preparation</li></ul>
Ontotext	<ul style="list-style-type: none"><li>• Architecture definition</li><li>• Selection of ontologies for the domain</li><li>• Queries definition</li><li>• User interface and integration?</li></ul>
UGOT	<ul style="list-style-type: none"><li>• Architecture definition</li><li>• Queries Grammar</li><li>• Patents Grammar</li></ul>

# Use Case Scenarios



# Basic Flow



# Queries topics

information about a drug	drugs that are compounds
active ingredients of a drug	drug preparations
dosage forms of a drug	the name of a drug
route of administration of a drug	methods in the patent
dosage form of a drug	use of patent
patent number	use of drug
the expiration of a patent	strength of a drug
patent use codes	claims from 2008 that mention a given drug
patent application number	claims about a given drug authored by somebody
applicant for a patent	
approval date of a patent	<b>missing from the ontology so far</b>
chemical substances of a drug/ pharmaceutical composition	diseases, illnesses

# Queries Examples (131 sentences)

what information can I get about A\_DRUG (aspirin)

what chemical substances there are in A\_DRUG?

what are the active ingredients of A\_DRUG (aspirin)

give me the drugs that are compounds

what are the dosage forms of A\_DRUG (aspirin)

the drug preparations for A\_DRUG with a patent that expires after DATE

what is the route of administration of A\_DRUG (aspirin)

I want the name of A\_DRUG with a patent with approval date DATE

what is the dosage form of A\_DRUG (aspirin)

what methods are used in THE\_PATENT?

what is the patent number of the patent for A\_DRUG

give me the use of patents approved in DATE / on DATE / before DATE / after DATE

when does THE\_PATENT expire?

give me the use codes of THE\_PATENT

# Grammar

- Patents Grammar
- Queries grammar
  - Full coverage of the examples and derivatives from them.

# SMT & the corpora horror story

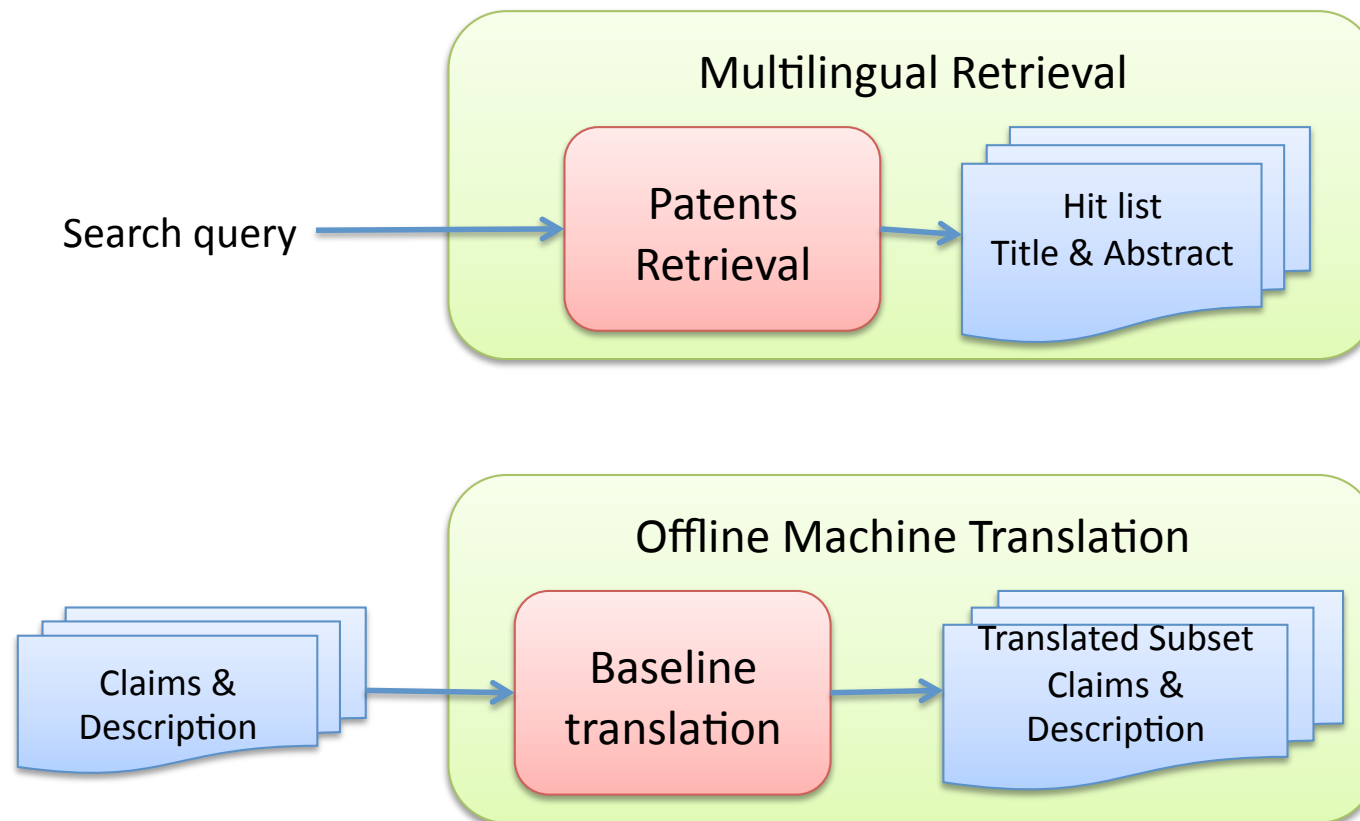
- Still, provisional version with parallel corpus extracted and prepared from provisional MAREC corpus with personal research license.
- We asked for WIPO corpus (and it has been shipped):
  - 8M segments aligned!
  - Only French and English
  - Only titles and abstracts
- Should we still wait for the EPO corpus?

# (near) Future work

D7.1 – Patent MT and Retrieval  
Prototype Beta



# Architecture Proposal



# GF & Ontology

- Translate the abstract representation of the queries to SPARQL.
- Transform the results from the semantic repository to ...

# GF & SMT

- Set up the baseline translation system: which?

# Tentative Calendar

May (M15)	June (M16)	July (M17)	August (M18)
	T7.1. User Req (WP9) T7.2. Corpora - MAREC Corpus	T7.7. Beta Prototype – Architecture	T7.2 Corpora - Compounds list and patents translated in the three languages
	T7.3. Grammar - preliminary report	T7.4 Ontology – Queries form	T7.3 Grammar - Queries grammar
September (M19)	October (M20)	November (M21)	December (M22)
D5.1. Corpora T7.3. Grammar - final report	T7.5 – PatRet - release a beta version T7.6. MT – realease a beta version	D7.1. Review user requirements by everybody	
T7.7. Prototype – Beta Interfaces and integration	T7.7. Prototype – Beta Testing	D7.1. Beta Prototype	
January (M23)	February (M24)	March (M25)	April (M26)
		D5.2. Hybrid prototype	

# Dissemination

- **Patent translation within the MOLTO project,**  
Cristina España-Bonet, Ramona Enache, Adam Slaskiy, Aarne Ranta, Lluís  
Màrquez & Meritxell Gonzalez,  
*MT Summit XIII - 4th Workshop on Patent Translation.*  
September 23, 2011 Xiamen, China