



**KYOTO** (ICT-211423) Intelligent Content and Semantics  
Knowledge Yielding Ontologies for Transition-Based Organization  
<http://www.kyoto-project.eu/>

Kybots, knowledge yielding robots  
German Rigau  
IXA group, UPV/EHU



ICT-211423

# KYOTO Overview

- **Title:** Knowledge Yielding Ontologies for Transition-Based Organization
- **Funded:**
  - 7<sup>th</sup> Framework Program-ICT of the European Union: Intelligent Content and Semantics
  - Taiwan and Japan funded by national grants
- **Goal:**
  - Platform for knowledge sharing across languages and cultures
  - Knowledge transition and information across different target groups, transgressing linguistic, cultural and geographic boundaries.
  - Open text mining and deep semantic search
  - Wiki environment that allows people in the field to maintain their knowledge and agree on meaning without knowledge engineering skills
- **URL:** <http://www.kyoto-project.eu/>
- **Duration:**
  - March 2008 – March 2011
- **Effort:**
  - 364 person months of work.



# KYOTO Overview

- **Languages:**
  - English, Dutch, Italian, Spanish, Basque, Chinese, Japanese
- **Domain:**
  - Environmental domain, BUT usable in any domain
- **Global:**
  - Both European and non-European languages
- **Available:**
  - Free: as open source system and data (GPL)
- **Future perspective:**
  - Content standardization that supports world wide communication
  - Global Wordnet Grid



# Consortium

1. Vrije Universiteit Amsterdam (Amsterdam, The Netherlands),
  2. Consiglio Nazionale delle Ricerche (Pisa, Italy),
  3. Berlin-Brandenburg Academy of Sciences and Humanities (Berlin, Germany),
  4. Euskal Herriko Unibertsitatea (San Sebastian, Spain),
  5. Academia Sinica (Tapei, Taiwan),
  6. National Institute of Information and Communications Technology (Kyoto, Japan),
  7. Irion Technologies (Delft, The Netherlands),
  8. Synthema (Rome, Italy),
  9. European Centre for Nature Conservation (Tilburg, The Netherlands),
- Subcontractors:
    - World Wide Fund for Nature (Zeist, The Netherlands),
    - Masaryk University (Brno, Czech)







# Ultimate goal

- Global standardisation and anchoring of meaning such that:
  - Machines can approach text understanding -> semantic web connects to the current web
  - Communities can dynamically maintain knowledge, concepts and their terms in an easy to use system
  - Cross-linguistic and cross-cultural sharing and communication of knowledge is enabled
  - Comparable to a formalization of Wikipedia for humans **AND** machines across languages

# Work Package List

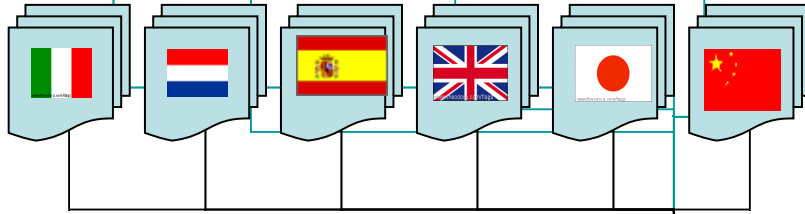
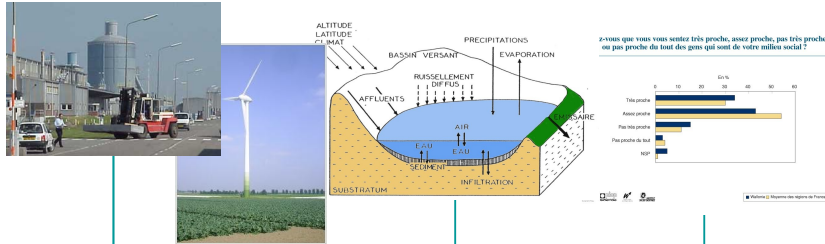
WP No	Work package title	Lead partic.	PM	Start	End
WP0	Management	VUA	9	1	36
WP1	User requirements	VUA	5	1	6
WP2	System design	SYNTHEMA	12	1	6
WP3	Capture	IRION	10	1	9
WP4	Indexing	IRION	11	4	12
WP5	Knowledge mining	EHU	120	7	30
WP6	Knowledge integration	BBAW	106	4	24
WP7	Database systems and wiki	CNR-ILC-IIT	25	1	24
WP8	Domain extension	ECNC	12	13	30
WP9	Evaluation	ECNC	20	4	33
WP10	Exploitation	SYNTHEMA	8	19	36
WP11	Dissemination	VUA	26	1	36
	TOTAL		364		



GREENPEACE



Environmental organizations



Citizens  
Governors  
Companies

Environmental organizations

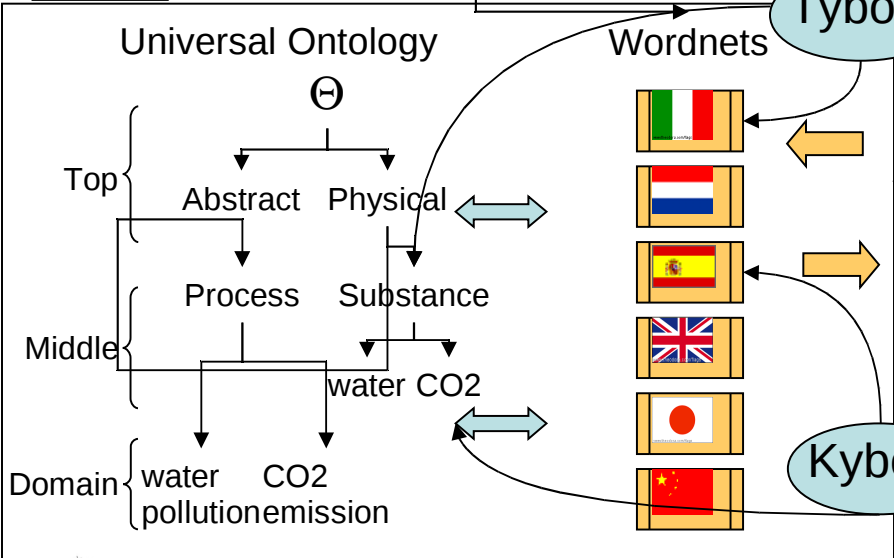


GREENPEACE

Domain Wikyoto

Global Wordnet Grid

Capture



Tybots

Wordnets

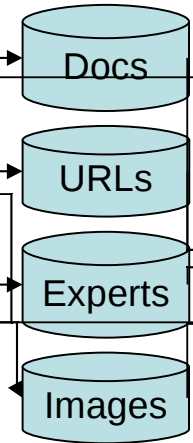


Kybots

Concept Mining

Fact Mining

Index



Dialogue  
Search

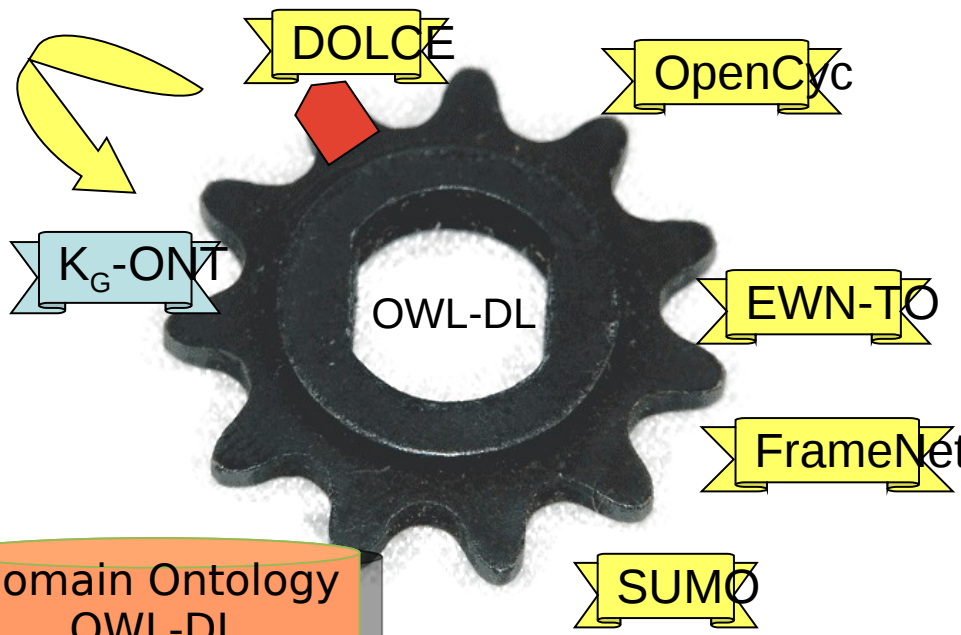


# System components

- Generic ontologies and databases
  - SUMO, DOLCE
  - Geo databases
  - Wikipedia
- Generic linguistic resources
  - Wordnet
  - FrameNet
- **Tybots**: Term yielding robots
- **Kybots**: knowledge yielding robots
- **Wikyoto**: wiki system for yielding domain wordnets and domain ontologies in social communities



Fact User



**Kybot Editor**

Kybot Profiles

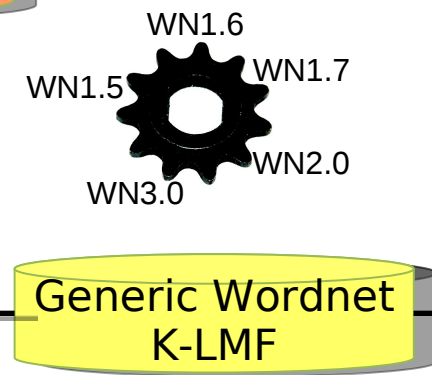
**Conceptual Pattern:**  
 causes, Process1, Process2  
 patient, Quantity, Increasing

**Kybot Server (Fact Extraction)**

Document Base KAF

ICT-211423

Domain Wordnet K-LMF → Domain Ontology OWL-DL



**Expression rules:**  
 -N+N  
 -N+prep+N  
 -Nsubj+V

**Term Editor (Wikyoto)**



Concept User

Extracted Terms Generic K-TMF



# KAF: Kyoto Annotation Framework

KAF is the input of both:

- Tybot: term extraction
  - Kybot: fact extraction
- 
- Word forms
  - Terms / items
  - Chunks
  - Dependencies
  - WSD / SRL
  - Events
  - Quantifiers
  - Time expressions
  - General Relations



# KAF word forms

- “John taught mathematics 20 minutes every Monday in New York.”

<text>

<wf wid="w1">John</wf>

<wf wid="w2">taught</wf>

<wf wid="w3">mathematics</wf>

<wf wid="w4">20</wf>

<wf wid="w5">minutes</wf>

<wf wid="w6">every</wf>

<wf wid="w7">Monday</wf>

<wf wid="w8">in</wf>

<wf wid="w9">New</wf>

<wf wid="w10">York</wf>

<wf wid="w11">.</wf>

</text>



# KAF terms

- “John taught mathematics 20 minutes every Monday in New York.”

<terms>

<term tid="t1" span="w1" type="entity" lemma="John" pos="N" netype="person"></term>

<term tid="t2" span="w2" type="open" lemma="teach" pos="V">

<senseAlt>

<sense sensecode="EN-17-00861095-v" weight="0.80"/>

<sense sensecode="EN-17-00859568-v" weight="0.20"/>

</senseAlt>

</term>

<term tid="t3" span="w3" type="open" lemma="mathematics" pos="N">

<senseAlt>

<sense sensecode="EN-17-04597590-n" weight="1.0"/>

</senseAlt>

</term>

<term tid="t4" span="w4" type="entity" lemma="20" pos="Z" netype="number"></term>

...



# KAF terms

...

```
<term tid="t5" span="w5" type="open" lemma="minute" pos="N"></term>
  <senseAlt>
    <sense sensecode="EN-17-12621100-n" weight="0.80"/>
    <sense sensecode="EN-17-12631889-n" weight="0.06"/>
    <sense sensecode="EN-17-12630443-n" weight="0.01"/>
    <sense sensecode="EN-17-11241911-n" weight="0.01"/>
    <sense sensecode="EN-17-05339359-n" weight="0.01"/>
    <sense sensecode="EN-17-04316149-n" weight="0.01"/>
  </senseAlt>
<term tid="t5" span="w6" type="close" lemma="every" pos="D"></term>
<term tid="t6" span="w7" type="entity" lemma="Monday" pos="N" netype="date"/>
  <senseAlt>
    <sense sensecode="EN-17-12557842-n" weight="1.0"/>
  </senseAlt>
<term tid="t7" span="w8" type="close" lemma="in" pos="P"></term>
<!-- multiword form -->
<term tid="t8" span="w9 w10" type="entity" lemma="New_York"
  pos="N"netype="location"></term>
</terms>
```



# KAF chunks

```
<chunks>  
  <!-- John -->  
  <chunk cid="c1" span="t1" head="t1" pos="NP"/>  
  <!-- mathematics -->  
  <chunk cid="c2" span="t3" head="t3" pos="NP"/>  
  <!-- in New York -->  
  <chunk cid="c3" span="t7 t8" head="t4" pos="PP"/>  
</chunks>
```



# KAF events

```
<events>
```

```
  <event eid="e1" span="t2" lemma="teach" pos="V" eiid="ei1" class="OCCURRENCE"  
    tense="PAST" aspect="NONE" polarity="POS">
```

```
  <roles>
```

```
    <role cid="c1" role="agent"/>
```

```
    <role cid="c2" role="subject"/>
```

```
    <role cid="c3" role="location"/>
```

```
  </roles>
```

```
</event>
```

```
</events>
```



# KAF quantifiers & time expressions

```
<!-- every -->
```

```
<quantifiers>
```

```
  <quantifier qid="q1" span="t5"/>
```

```
</quantifiers>
```

```
<!-- 20 minutes every monday -->
```

```
<timexs>
```

```
  <timex3 texid="tex1" span="t4 t5" type="DURATION" value="P20TM"/>
```

```
  <timex3 texid="tex2" span="t5 t6" type="SET" value="xxxx-wxx-1"  
    quant="EVERY"/>
```

```
  <tlink timeID="tex1" relatedToTime="tex2" relType="IS_INCLUDED"/>
```

```
  <tlink eventInstanceID="ei1" relatedToTime="tex1" relType="SIMULTANEOUS"/>
```

```
</timexs>
```



# What Tybots do...

- Input are text documents
- Linguistic processors generate KAF annotation:
  - morpho-syntactic analysis
  - semantic roles
  - named entities
  - wordnet and ontology mappings
- Output are term hierarchies in TMF:
  - structural parent relations
  - quantified structural and semantic relations
  - statistical data
  - generalized semantic mappings



Bio:0.82#Fishing:0.8# Tops:206, Nodes:476, Docs:397

Import View Edit Save Sort Clear

- external threats:2:1::3
  - Managing External Threats:2:1::1
  - key external threats:1:1::1
  - threats of oil extraction:1:1::1
- reef:0::10
  - largest known cold water coral reef:1:1::1
  - important reefs:1:1::1
  - Great Barrier Reef:1:1::1
  - Barrier Reef:6:2::2
  - coral reef:4:1::5
    - unique coral reefs:1:1::1
    - difficult coral reef:2:1::1
    - water coral reef:2:1::2
      - cold water coral reef:2:1::1
- habitats:1:1::9
  - sensitive habitats:2:1::2
    - harm sensitive habitats:2:1::1
  - affecting habitats:2:1::3
    - activities affecting habitats:2:1::2
      - human activities affecting habitats:2:1::1
  - marine habitats:2:1::2
    - sensitive marine habitats:1:1::1
  - multi-habitat:1:1::1
- species:8:2::7
  - fish species:2:1::2
    - key fish species:1:1::1
  - multi-species:2:1::1
  - one-species:2:1::1
  - marine species:2:1::1

Done loading thesaurus...

Lema1

Lema2

Lema3

Lema4

Hautagaien ordena Luzeera

- Alfabetoa  1
- Maiztasuna  2
- Estatistika  3
- ZT hiztegiak  4
- Euskalterm bankukoak

Erauzketa hautatu

Ekologia

Aeron.

bidali  esportatu

Forma	Eredu	Alor.	Maiz.	Neur.	Test.	Ter.?
eskuinaldeko ertz ↑ ↓				1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
haltzadi kantauriarra ↑ ↓		Aeron.	(186)	1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
harizti-baso misto ↑			129	1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
ezkerraldeko ertz ↑ ↓		Aeron.	(364)	1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
ibaiaren arro ↑ ↓ M	AprepN	Aeron.	176 (14)	1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
harizti-baso ↑ ↓	NN	Aeron.	130 (130)	1000.00	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
jarduera poluitzaile ↓			101 (101)	763.59	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
balio limite ↑ ↓			99 (79)	752.83	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
B mota ↑ ↓	NN	Aeron.	107 (107)	621.12	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
B motako arau ↑	NAprepN	Aeron.	103 (103)	613.84	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
hurrengo orrialde ↑	AprepN	Aeron.	99 (53)	611.28	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
ertz-ekerralde ↓	NN	Aeron.	54 (54)	515.65	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
harizti azido ↑	NApos	Aeron.	55 (54)	511.84	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>
ikas-norabide ↑ ↓	NN	Aeron.	55 (55)	501.08	<a href="#">KWIC</a>	<input checked="" type="checkbox"/> <input type="checkbox"/>

Basque Mountain range (?)

Oaktree mixed forest (?)

Polluting activities (?)

# Infomap BNC + SSI-Dijkstra

associate -n 20 -c BNCpos3prova "tropicalpa" "speciespn"

**tropical**|a 0.953014

**species**|n 0.953014

birds|n 0.926641

mammals|n 0.908901

invertebrates|n 0.889433

breeding|n 0.881263

temperate|a 0.876306

prey|n 0.873921

bird|n 0.869077

whales|n 0.865983

insects|n 0.861247

habitat|n 0.854986

predators|n 0.853619

butterflies|n 0.845556

frogs|n 0.827578

genus|n 0.827000

fauna|n 0.822362

arctic|a 0.821317

habitats|n 0.820968

seals|n 0.818886

animals|n 0.815580

...



# Infomap + SSI-Dijkstra

[rigau@adimen MCRGraphDistances]\$ ./SSI-Dijkstra-en30.pl

Reading Graph from file ...

Polysemous: **tropical**|a 4

Polysemous: **species**|n 2

Polysemous: breeding|n 5

Polysemous: temperate|a 3

Polysemous: prey|n 2

Polysemous: bird|n 5

Monosemous: habitat|n 1

Polysemous: genus|n 2

Polysemous: fauna|n 2

Interpretation: breeding n 00914929-n 0.464285714285714 7 the production of animals or plants by inbreeding or hybridization

Interpretation: fauna n 00015388-n 0.5 1 a living organism characterized by voluntary movement

Interpretation: temperate a 02402559-a 0.383333333333333 5 (of weather or climate) free from extremes; mild; or characteristic of such weather or climate

Interpretation: habitat n 08580583-n 0 0 the type of environment in which an organism or group normally lives or occurs

Interpretation: bird n 01503061-n 0.4375 8 warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings

Interpretation: **species** n 08110373-n 0.416666666666667 2 (biology) taxonomic group whose members can interbreed

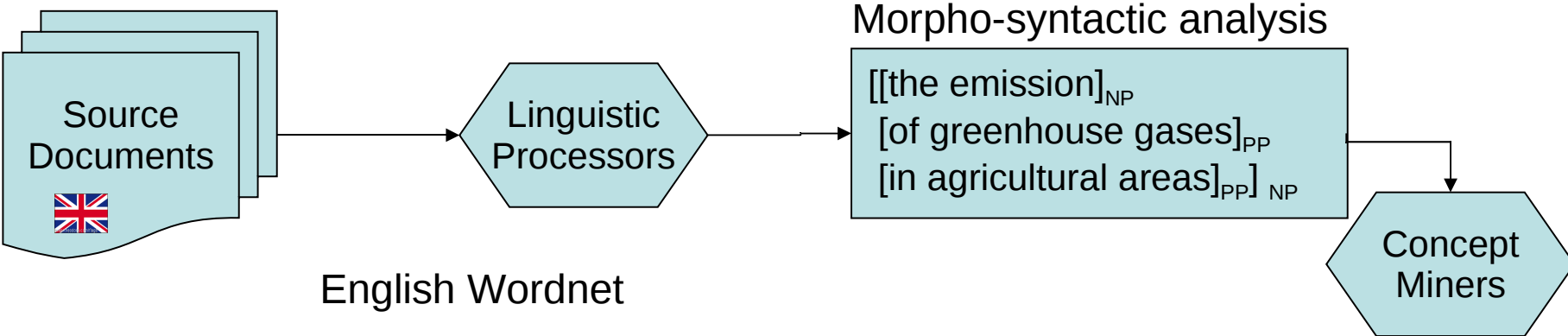
Interpretation: **tropical** a 02443907-a 0.347222222222222 6 relating to or situated in or characteristic of the tropics (the region on either side of the equator)

Interpretation: prey n 02152881-n 0.555555555555555 3 animal hunted or caught for food

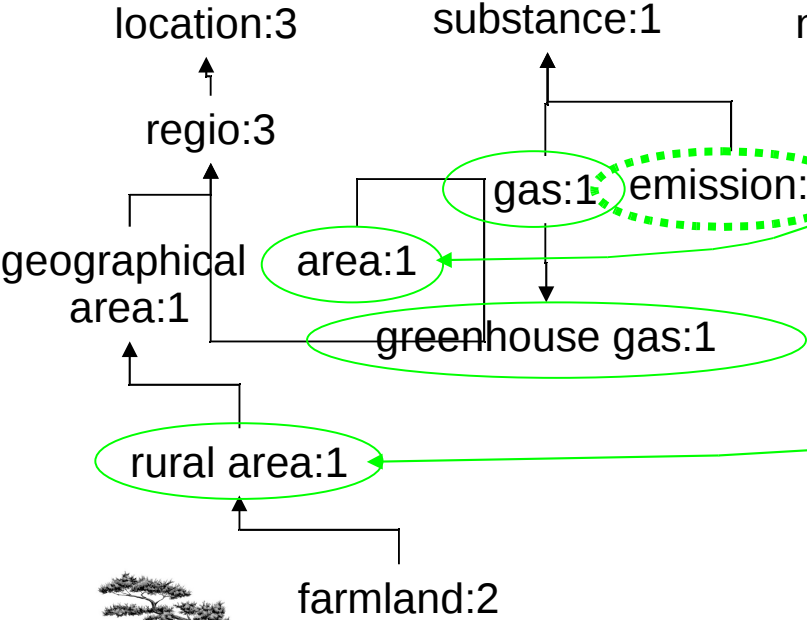
Interpretation: genus n 08108972-n 0.583333333333333 4 (biology) taxonomic group containing one or more species



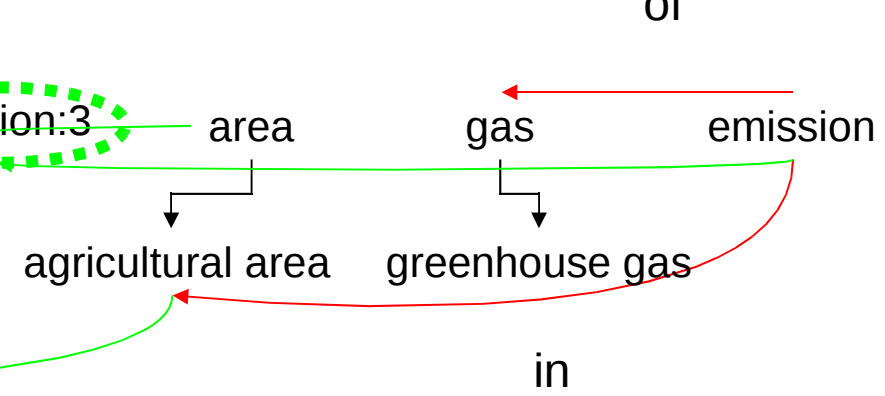
# Concept mining by Tybots



English Wordnet



Term hierarchy





# Kybots, knowledge yielding robots

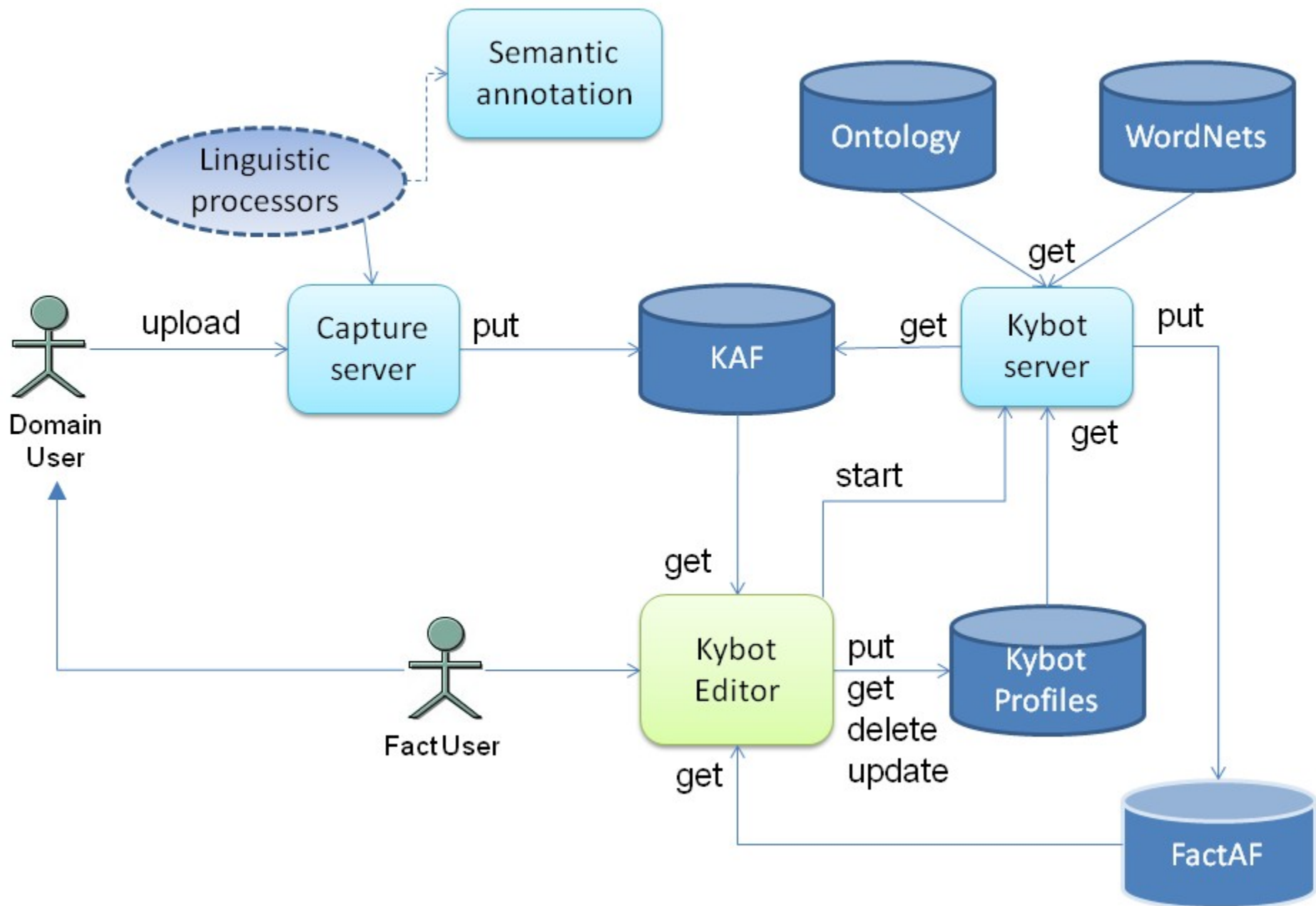
- What kybots do?
- Building Kybots:
  - Mining by example
- KAF
- Kybot profiles
- Kybots



# Knowledge Mining

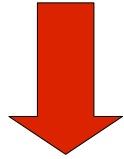
- Concept mining (Tybot)
  - Extract terms and relations in a language
  - Map the terms to an existing wordnet
  - Ontologize terms to concepts and axioms
- Fact mining (Kybot)
  - Enrich text with linguistic and semantic information
  - Define patterns in text
  - Extract facts from text
- For all languages!

# KYBOT EDITOR

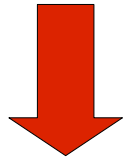
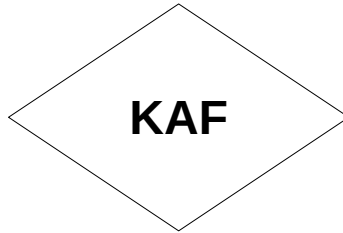


# Fact Mining: Building kybots

Tropical terrestrial species populations declined by 55 per cent on average from 1970 to 2003



- + Linguistic Processing: POS, chunks, dependencies, ...
- + Semantic Processing: WSD, SRL, ... (WN, ontologies)



- + Kybot Editor: Kybot profiles
- + Kybot Engine: FactAF

Tropical terrestrial species **populations declined** by **55** per cent on average from 1970 to 2003

# Building Kybots: Mining by example

- Kybots perform a complex **Information Extraction (IE)** task requiring expertise on:
  - linguistic engineering
  - knowledge engineering ...
- but ...
  - all this complexity should be hidden to the end-user
- Our proposal is to build complex kybots using an advanced wiki system following a new approach:
  - Mining by example

# Building Kybots: Mining by example

- Kybot editor allows to **mine by example** the domain corpus for helping users to define Kybot profiles
- Users define kybots of their interest ...
  - Input:
    - a collection of captured **domain documents**
    - a set of **information needs** or **questions**
    - a set of textual snippets which **support** the answers to the questions
  - Output:
    - a collection of Kybot profiles



# Building Kybots: Mining by example

- a) Use a **basic IR system** consulting the domain corpus.
  - input: "population decline", "decrease population", ...
  
- b) Inspecting the resulting snippets.
  
- c) A kybot profile is defined selecting the **relevant information** from each snippet
  - how many, where, when, ...
  
- d) Kybots are applied on the document collection.
  - Kybots use all the capabilities of the linguistic processors, including domain wordnet, general wordnets, ontologies, reasoning, etc.



# Building Kybots: Mining by example

- information need:
  - “reduction of populations”
- Looking for answers to the following questions:
  - Which species?
  - Degree of the reduction?
  - Period of time?
- Textual snippet supporting the answers:
  - “Tropical terrestrial species populations declined by 55 percent on average from 1970 to 2003”
- Resulting Kybot profile:
  - `kybot_decrease_of_population`





# Building Kybots: Mining by example


- “Tropical terrestrial species populations **declined** by 55 per cent on average from 1970 to 2003”
- **declined** is enriched now with KAF information:
  - Word form: “declined”
  - Part-of-speech: Verb
  - Lemma: “decline”
  - Linguistic references to other elements in text ...
  - Ranked list of senses:
    - 00149782-v worsen\_1 decline\_1 grow worse: ...
    - 01670714-v refuse\_2 reject\_2 pass\_up\_1 turn\_down\_1 decline\_2 refuse to accept: ...
    - 00587372-v refuse\_1 decline\_3 show unwillingness towards: ...
    - 01520410-v decline\_5 go down: ...
    - ...
  - Wordnet information: general concepts, BC, ...
  - Ontological information: DOLCE, SUMO, TCO, ...

**Search word**

Select Collection : english

Word: decrease Search

**Found 8 sentences with term or word equal to "decrease"**

 /db/kyoto/kaf2/english/490.kaf2

This has accompanied a continuing decrease in the importance of farming

Example: term: decrease - type: open - POS: n

TERM:	decrease
POS:	n
TYPE:	open

**Sense**

- [0.274686] decrease, lessening, drop-off : a change downward
- [0.261167] decrease, diminution, reduction, step-down : the act of decreasing or reducing something
- [0.233946] decrease, decrement : a process of becoming smaller or shorter
- [0.2302] decrease, decrement : the amount by which something decreases

areas where precipitation increases in spring decreases summer

http://xmlgroup.iit.cnr.it/cocoon/kybot/index.xql





# Wikybot

- [Home](#)
- [Upload documents](#)
- [Manage kybots](#)
- [Manage snippets](#)

Hello rigau!

Logout

Click on a text portion of a snippet, then select the role to associate. Click on a labelled snippet text portion to delete its role association.

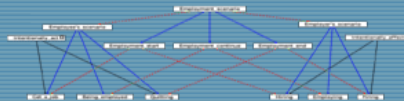
Role colors : Agent Patient Location Time Topic

Document : *Living planet english*

--- The other index in this report, the *Living Planet Index*, shows a rapid and continuing loss of biodiversity - populations of vertebrate species have declined by about one third since 1970.

--- populations of terrestrial species declined by about 30 per cent on average between 1970 and 2003.

--- The rapid rate of population decline in tropical species is mirrored by the loss of natural habitat to cropland or pasture in the tropics between 1950 and 1990 , agricultural conversion being the main driver.




[Main Menu](#)
[Home](#)
[View FrameNet Data](#)
[News](#)
[Licensing FrameNet](#)
[Data](#)
[FAQs](#)
[The Book \(PDF\)](#)
[Presentations and](#)
[Talks](#)
[FrameNet Related](#)
[Projects](#)
[Publications and](#)
[Papers](#)
[FN2 Forum](#)
[Staff](#)
[Users](#)
[Funding](#)
[Current Project](#)
[Status](#)
[Done](#)

## Frame Report (recent data)

[| Top of Frame Index](#) | [| Top of Lexical Unit Index](#) |

# Change\_position\_on\_a\_scale

### Definition:

This frame consists of words that indicate the change of an **Item**'s position on a scale (the **Attribute**) from a starting point (**Initial\_value**) to an end point (**Final\_value**). The direction (**Path**) of the movement can be indicated as well as the magnitude of the change (**Difference**). The rate of change of the value (**Speed**) is optionally indicated. Another scale (**Correlate**), which the values are correlated with, is indicated if it is not the default correlate (namely, absolute time).

The distinction between **Attribute**s and **Items** is not always an easy one. The clear cases involve the expression of the **Attribute** in an *in*-PP.

**Hawke's Bay winery** DOUBLED **in size** **last year**.

The amount you can deduct, then, depends on whether or not **the stock** has INCREASED **in value** **during the period you have owned it**.

Other clear cases of **Attribute**s involve NP's like *size, quality, number, value* that denote abstract attributes rather than events or classes of things.

The analysis is more complicated when an event-denoting noun phrase occurs as the subject and the sentence has no *in*-PP. In the simplest cases, where the **Attribute** is the number of occurrences of the event, the **Attribute** is usually left implicit, as in exx. 1 and 2, the **Attribute** below.

(1) **Accidents** INCREASED **20% to 345**.

(1--second FE layer) **Accidents** INCREASED **20% to 345**.

(2) **Attacks on civilians** DECREASED **over the last 4 months** **DNI**.

- [The Book \(PDF\)](#)
- [Presentations and Talks](#)
- [FrameNet Related Projects](#)
- [Publications and Papers](#)
- [FN2 Forum](#)
- [Staff](#)
- [Users](#)
- [Funding](#)
- [Current Project](#)
- [Status](#)

## Frame Elements and Their Syntactic Realizations

The Frame elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realizations(s)
Attribute	(52)	2nd.-- (2) DNI.-- (1) INI.-- (14) NP.Ext (24) PP[in].Dep (11)
Difference	(42)	INI.-- (32) AVP.Dep (4) PP[by].Dep (5) NP.Obj (1)
Final_value	(11)	PP[about].Dep (1) PP[to].Dep (10)
Initial_value	(6)	PP[from].Dep (6)
Item	(52)	NP.Ext (28) DNI.-- (16) 2nd.-- (8)

### Valence Patterns:

These frame elements occur in the following syntactic patterns:

Number Annotated	Patterns				
1 TOTAL	Attribute	Difference	Final_value	Initial_value	Item
(1)	NP Ext	PP[by] Dep	PP[to] Dep	PP[from] Dep	DNI --
2 TOTAL	Attribute	Difference	Final_value	Item	
(1)	NP Ext	NP Obj	PP[to] Dep	DNI --	
(1)	NP Ext	PP[by] Dep	PP[to] Dep	DNI --	
39 TOTAL	Attribute	Difference	Item		
(1)	DNI --	INI --	NP Ext		
(2)	INI --	AVP Dep	NP Ext		

- [Contact Us](#)
- [Template by Mambo](#)
- [Hut](#)

Login Form

Username

Password

Remember me

Login

[Password Reminder](#)

[No account yet?](#)

[Create one](#)

Applications Places System 0°F German Rigau Tue Feb 3, 11:44

Web MCR Interface MEANING - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl?item=decline&button1=Look\_up&metode=Word&pos=Verbs&llengua=English\_1.6&search=n

Most Visited Release Notes Fedora Project Red Hat Free Content

decline Look up

Word Verbs English\_1.6

near\_synonym English\_1.6

Gloss  English\_1.6  Italian\_1.6  Catalan\_1.5  English\_3.0

Score  Spanish\_1.6  English\_1.7  Spanish\_1.5

Rels  Catalan\_1.6  English\_1.7.1  English\_1.5

Full  Basque\_1.6  English\_2.0  English\_2.1

**Multilingual Central Repository**

---

**00139555v**  
[medicine](#) 00139555v 25 [worsen\\_1 decline\\_1](#) *grow worse: Conditions in the slum worsened;*

**base concept** 00139555v 24 [agravar\\_1 empeorar\\_1 decaer\\_1](#)

**change** 00139555v 26 [gainbehera\\_etorri\\_1 okerragotu\\_1 okerrera\\_egin\\_1 txarrera\\_egin\\_1 txartu\\_1](#)

**Decreasing BoundedEvent Condition Dynamic** 00139555v 0 [acuirsi\\_1 peggiorare\\_1](#) *grow worse: Conditions in the slum worsened;*

00203866v 28 [worsen\\_1 decline\\_1](#)

---

**01531148v**  
[factotum](#) 01531148v 2 [refuse\\_2 reject\\_2 pass\\_up\\_1 turn\\_down\\_1 decline\\_2](#) *refuse to accept: He refused my offer of hospitality;*

**base concept** 01531148v 2 [declinar\\_2 rechazar\\_1](#)

**possession** 01531148v 3 [-i\\_ezetz\\_eman\\_1 -i\\_uko\\_egin\\_1 baztertu\\_14](#)

**Committing** 01531148v 3 [recusare\\_2 respingere\\_2 ricasare\\_1 rifiutare\\_1 rigettare\\_1](#)

**SecondOrderEntity** 02237338v 3 [refuse\\_2 reject\\_2 pass\\_up\\_1 turn\\_down\\_1 decline\\_2](#) *refuse to accept: He refused my offer of hospitality;*

---

**00541774v**  
[factotum](#) 00541774v 8 [refuse\\_1 decline\\_3](#) *show unwillingness towards*

**base concept** 00541774v 8 [declinar\\_1 rehusar\\_1](#)

**communication** 00541774v 8

Done

rigau resea... proje... kyoto Gmail ... meeti... 2009-... kybot... Web ...

http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl



# Building Kybots: Mining by example

- “Tropical terrestrial species populations declined by 55 per cent on average from 1970 to 2003”
- A Wiki system will allow users to select/edit KAF information for building kybot profiles
  - general linguistic and semantic patterns
- For instance: kybot\_decrease\_of\_population
  - Looking for the degree of decrement:
    - 55%
    - 75 percent
    - ...
  - when it is a decrement of population ...
    - decline, worsen, ...
    - concepts, more general concepts ...
    - The class of verb of change followed by preposition followed by...
    - ...



# Linguistic Processors

- KAF (Kyoto Annotation Format)
  - English: **Synthema**
  - Dutch: **VUA**
  - Italian: **Synthema**
  - Basque: **EHU**
  - Spanish: **EHU**
  - Chinese: **AS**
  - Japanese: **NICT**



# Linguistic Processors

- KAF (Kyoto Annotation Format) is the input of both:
  - Tybot: term extraction
  - Kybot: fact extraction
- XML files including sections for:
  - Word forms
  - Terms / Items
  - Chunks: grouping of sequences of terms
  - Dependencies: syntactic relations between terms
  - WSD: senses of the term
  - SRL: roles of the term
  - Events
  - Quantifiers
  - Time expressions
  - General Relations
  - ...



# Kybot engine

- Kybots:
  - Process analysed text (KAF input)
  - Generates facts (factAF output)
- Kybot profiles
  - Expression Rules
    - Conditions on the LPs outcomes
    - Flexible enough for dealing with all KAF outputs
    - Capture **info** from the input
  - Semantic Conditions: WordNets + Ontologies
    - Semantic conditions on the **info**
    - Inferencing on the ontology / WN
  - Output Template
    - Expression consistent with the ontology



# Kybot engine

- Applies the Kybot profile to the analysed text (KAF file)
  - Subtrees of the Expression Rules (XPath-like)
  - Semantic Conditions (DEB API calls)
  - Output Templates (FactAF)
- For each analysed sentence :
  - **IF**
    - Expression Rules match **and**
    - Semantic Conditions hold
  - **THEN**
    - generate the Output Template

# Compiling kybot profiles and running kybots

- Kybots are described by **Kybot profiles** and once compiled they become XSLT / XQUERY scripts
- XSLT / XQUERY scripts can process KAF files
- Compiling a **kybot profile** (XSLT)
  - % ./kybotc kybot\_profile\_001.kybot > kybot001.xsl
- Running a **Kybot** (XSLT)
  - % xsltproc kybot001.xsl 2708\_sense.kaf.xml

# Kybot profiles: XSLT example 1

```
# KYBOT-PROFILE-QUANTITY-CHANGE-0001
```

```
# ... decrease by Z% ...
```

```
terms:
```

```
$V=term(@pos="v*"&sense(@sensecode="00111597-v"))
```

```
..1
```

```
$P=term(@pos="p*")
```

```
..1
```

```
$Z=term(@lemma="*%"|@lemma="*percent")
```

```
fact:
```

```
fact_name="quantity-change-001"
```

```
"term"=$Z(@tid)
```

```
"quantity"=$Z(@lemma)
```

# Running Kybots

```
% xsltproc kybot001.xsl 2708_sense.kaf.xml
<?xml version="1.0"?>
<fact id="quantity-change-001">
  <factval name="term" value="t4688"/>
  <factval name="quantity" value="30 percent"/>
</fact>
<fact id="quantity-change-001">
  <factval name="term" value="t4843"/>
  <factval name="quantity" value="r5.5 percent"/>
</fact>
```



# Kybot profiles: XQUERY example 1

**query** '

```
let $V := collection("proba.dbxml")//term[starts-with(@pos,"v")
and ./sense[@sensecode="00111597-v"]]
```

```
let $P := $V/following-sibling::node()[1][starts-with(@pos,"p")]
```

```
let $Z := $P/following-sibling::node()[1][ends-with(@lemma,"%") or
ends-with(@lemma,"percent")]
```

**return**

```
<fact id="quantity-change-001">
```

```
<factval name="terms" value="{ $Z/@tid }"/>
```

```
<factval name="quantitys" value="{ $Z/@lemma }"/>
```

```
</fact>'
```

# Kybot profiles: XSLT example 2

KYBOT-PROFILE-QUANTITY-CHANGE-0002

# ... decrease by Z% ...

terms:

\$V=term(@lemma="decrease" | @lemma="increase")

..1

\$P=term(@pos="p\*")

..1

\$Z=term(@lemma="\*%" | @lemma="\*percent")

fact:

fact\_name="quantity-change-002"

"term"=\$Z(@tid)

"quantity"=\$Z(@lemma)



# Kybot profiles: XQUERY example 2

**query** '

```
let $V := collection("proba.dbxml")//term[@lemma="decrease" or  
@lemma="increase"]
```

```
let $P := $V/following-sibling::node()[1][starts-with(@pos,"p")]
```

```
let $Z := $P/following-sibling::node()[1][ends-with(@lemma,"%") or  
ends-with(@lemma,"percent")]
```

**return**

```
<fact id="quantity-change-001">
```

```
<factval name="terms" value="{ $Z/@tid }"/>
```

```
<factval name="quantitys" value="{ $Z/@lemma }"/>
```

```
</fact>'
```

# Current issues

- Expressivity of the Kybot profiles
  - Focussing on Terms:
    - Complex expressions
    - External functions to access WN + ontologies
    - Variable scope “...” inside a sentence
    - Changing focus with parenthesis
    - Regular expressions on attributes
    - ...

# Open issues

- Expressivity of the Kybot profiles
  - Focussing on Dependencies ...
  - Focusing on Chunks ...
  - Combination of terms/dependencies/chunks
  - Output templates / KAF transformations
  - ...
- Running kybots
  - XSLT / XQUERY scripts
  - Efficiency vs. expressivity
  - Internal KAF representation for efficiency / indexing
  - Combination of kybots
  - ...



**KYOTO** (ICT-211423) Intelligent Content and Semantics  
Knowledge Yielding Ontologies for Transition-Based Organization  
<http://www.kyoto-project.eu/>

Kybots, knowledge yielding robots

German Rigau

IXA group, UPV/EHU

First Review Meeting

March 17, 2009, Luxembourg



ICT-211423