KYOTO (ICT-211423) Intelligent Content and Semantics
Knowledge **Y**ielding **O**ntologies for **T**ransition-Based **O**rganization
http://www.kyoto-project.eu/

Kybots, knowledge yielding robots

German Rigau
IXA group, UPV/EHU

KYOTO Overview

- **Title: K**nowledge **Y**ielding **O**ntologies for **T**ransition-Based **O**rganization
- **Funded:**
  - 7th 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 Humantities (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),

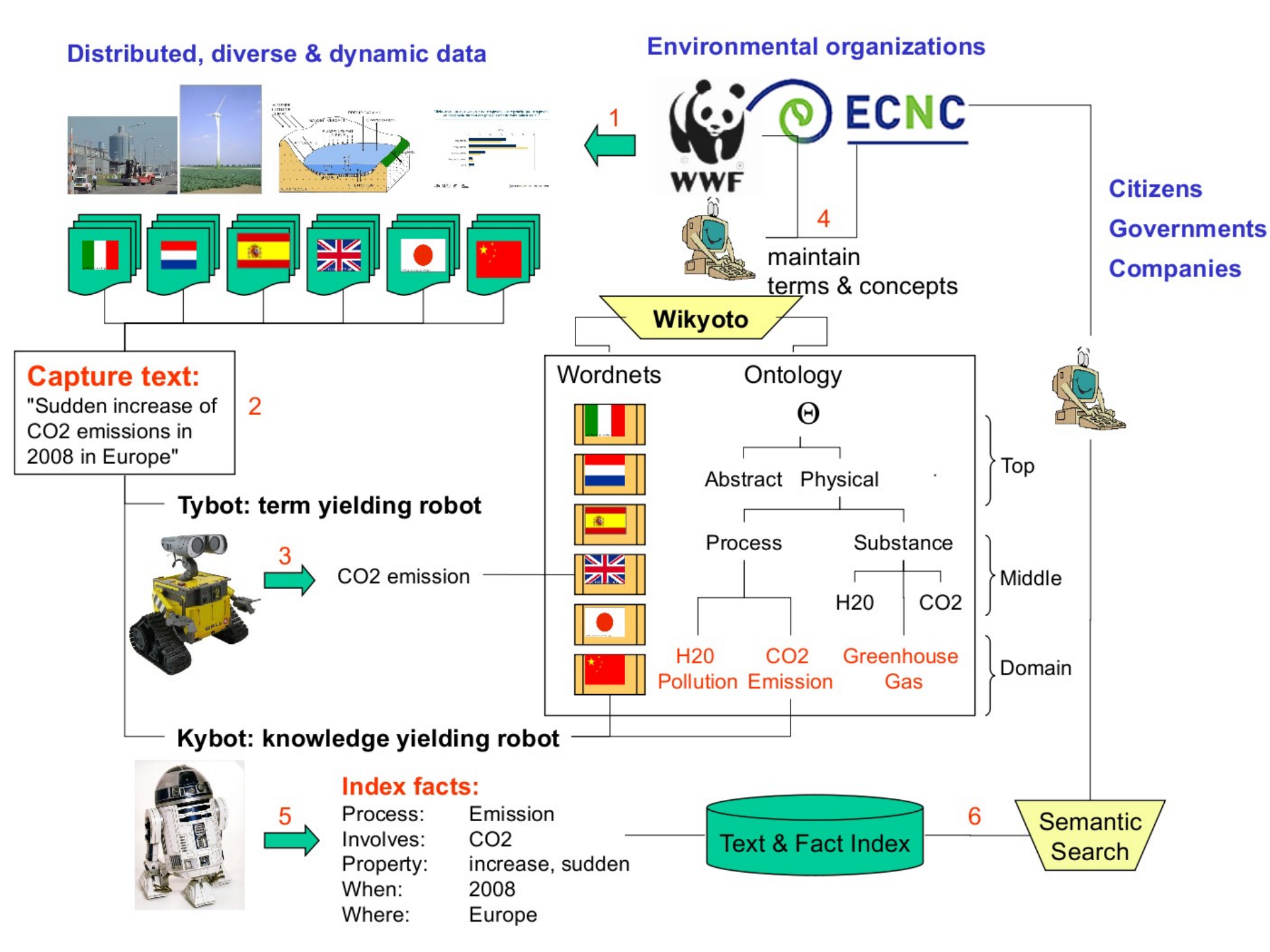
- Subcontractors:
  - World Wide Fund for Nature (Zeist, The Netherlands),
  - Masaryk University (Brno, Czech)

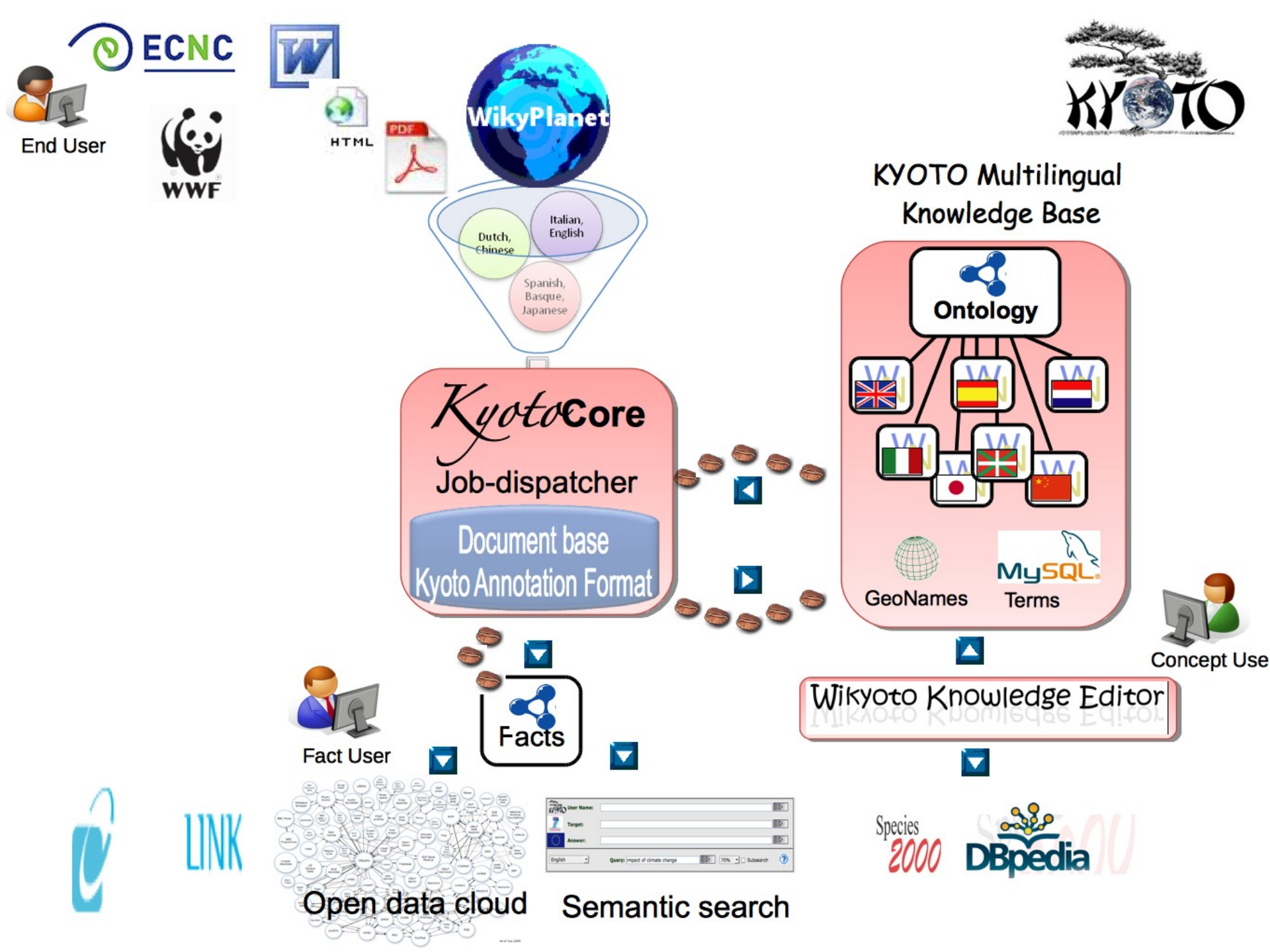
Ultimate goal

- 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		


Distributed, diverse & dynamic data
Environmental organizations
1
WWF
ECNC
4
maintain
terms & concepts
Citizens
Governments
Companies
Wikyoto
Capture text:
"Sudden increase of CO2 emissions in 2008 in Europe"
2
Wordnets
Ontology
Θ
Abstract
Physical
Process
Substance
H20
CO2
H20 Pollution
CO2 Emission
Greenhouse Gas
Top
Middle
Domain
Tybot: term yielding robot
3
CO2 emission
Kybot: knowledge yielding robot
5
Index facts:
Process: Emission
Involves: CO2
Property: increase, sudden
When: 2008
Where: Europe
Text & Fact Index
6
Semantic Search


ECNC
End User
WWF
HTML
PDF
WikyPlanet
Dutch, Chinese
Italian, English
Spanish, Basque, Japanese
KYOTO
KYOTO Multilingual Knowledge Base
KyotoCore
Job-dispatcher
Document base
Kyoto Annotation Format
Ontology
GeoNames
MySQL
Terms
Concept User
Wikyoto Knowledge Editor
Facts
Fact User
LINK
Open data cloud
Semantic search
Species 2000
DBpedia

KyotoCore

Capture

Html

Document base

Job dispatcher

PipeT

Modules

LP-client

MW-tagger

Sense-tagger^{UKB}

NE-tagger

ON-tagger

Tybot

Kybot

LP-server

MW-tagger
Sense-tagger
NE-tagger
ON-tagger
Kybot

KAF/lp

KAF DB

KAF DB

KAF/lp

KAF/ont

KAF/ont

LP-server

MW-tagger
Sense-tagger
NE-tagger
ON-tagger
Tybot

GeoNames

Knowledge Repository

Profiles

Facts

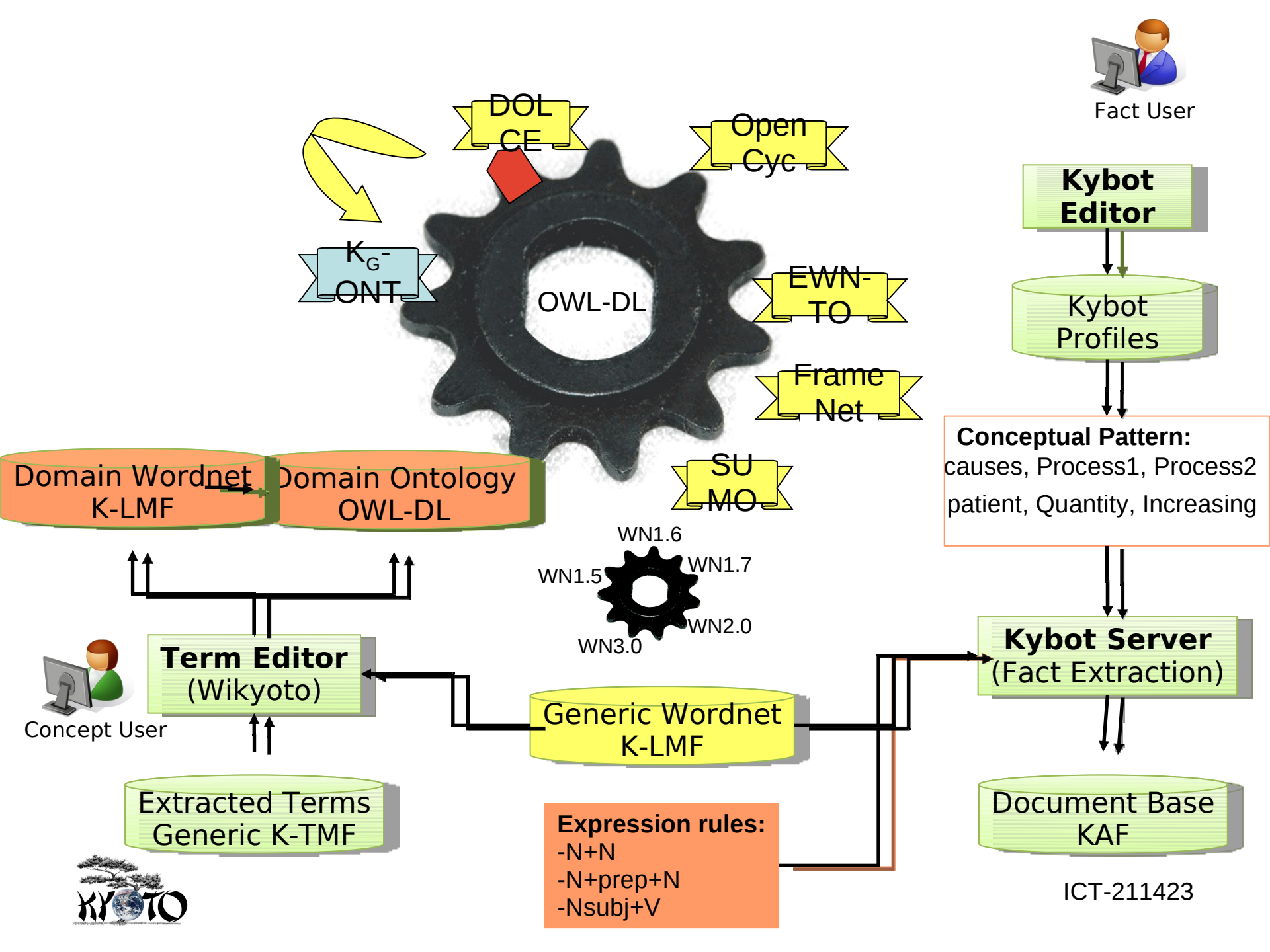
W terms

MySQL

Wikyoto Knowledge Editor

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
DOLCE
Open Cyc
K_G-ONT
OWL-DL
EWN-TO
Frame Net
SUMO
Kybot Editor
Kybot Profiles
Conceptual Pattern:
causes, Process1, Process2
patient, Quantity, Increasing
Domain Wordnet K-LMF
Domain Ontology OWL-DL
WN1.6
WN1.5
WN1.7
WN2.0
WN3.0
Term Editor (Wikyoto)
Concept User
Generic Wordnet K-LMF
Kybot Server (Fact Extraction)
Extracted Terms Generic K-TMF
Expression rules:
-N+N
-N+prep+N
-Nsubj+V
Document Base KAF
ICT-211423

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>
```

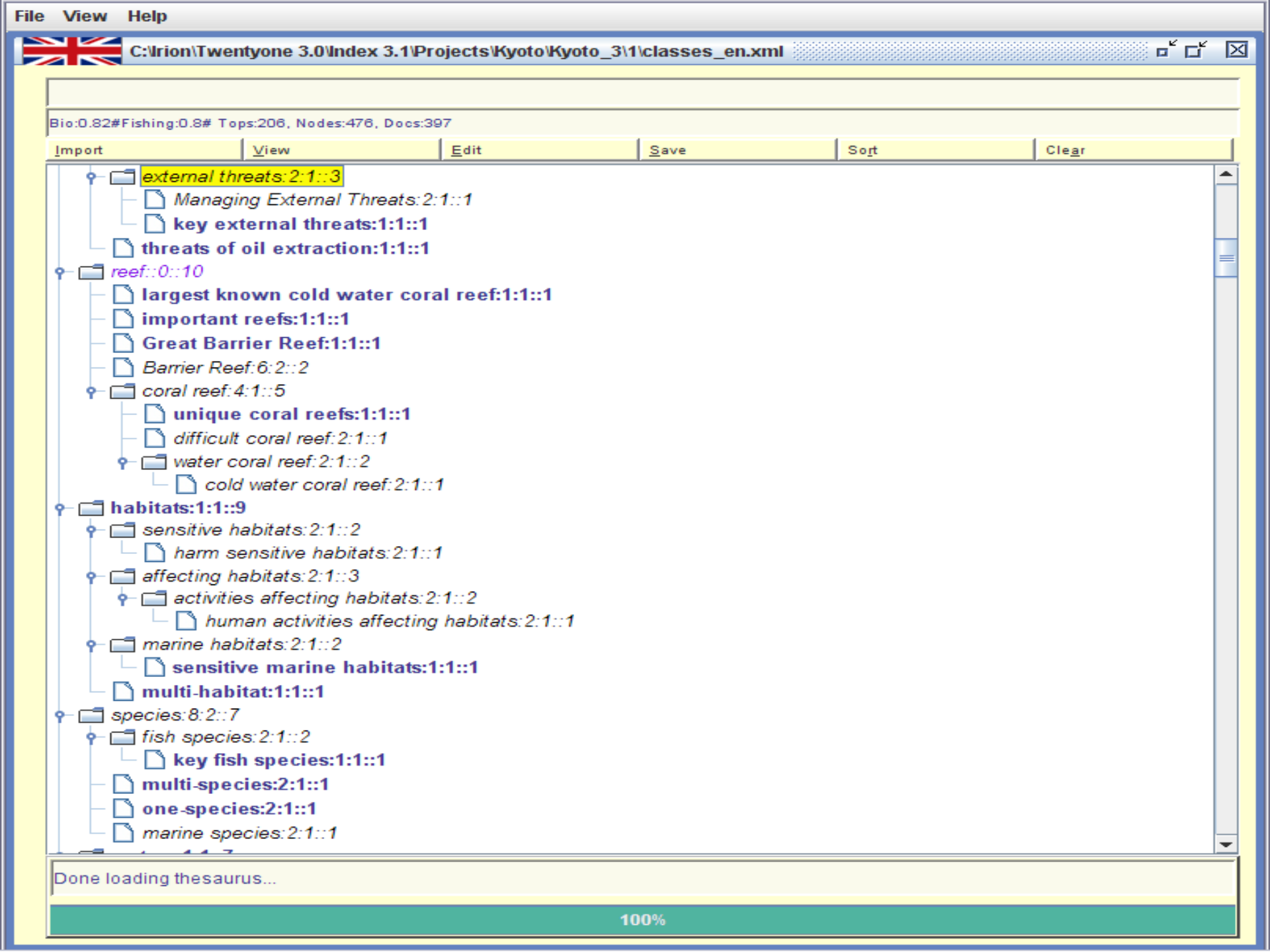
KAF quantifiers & time expressions

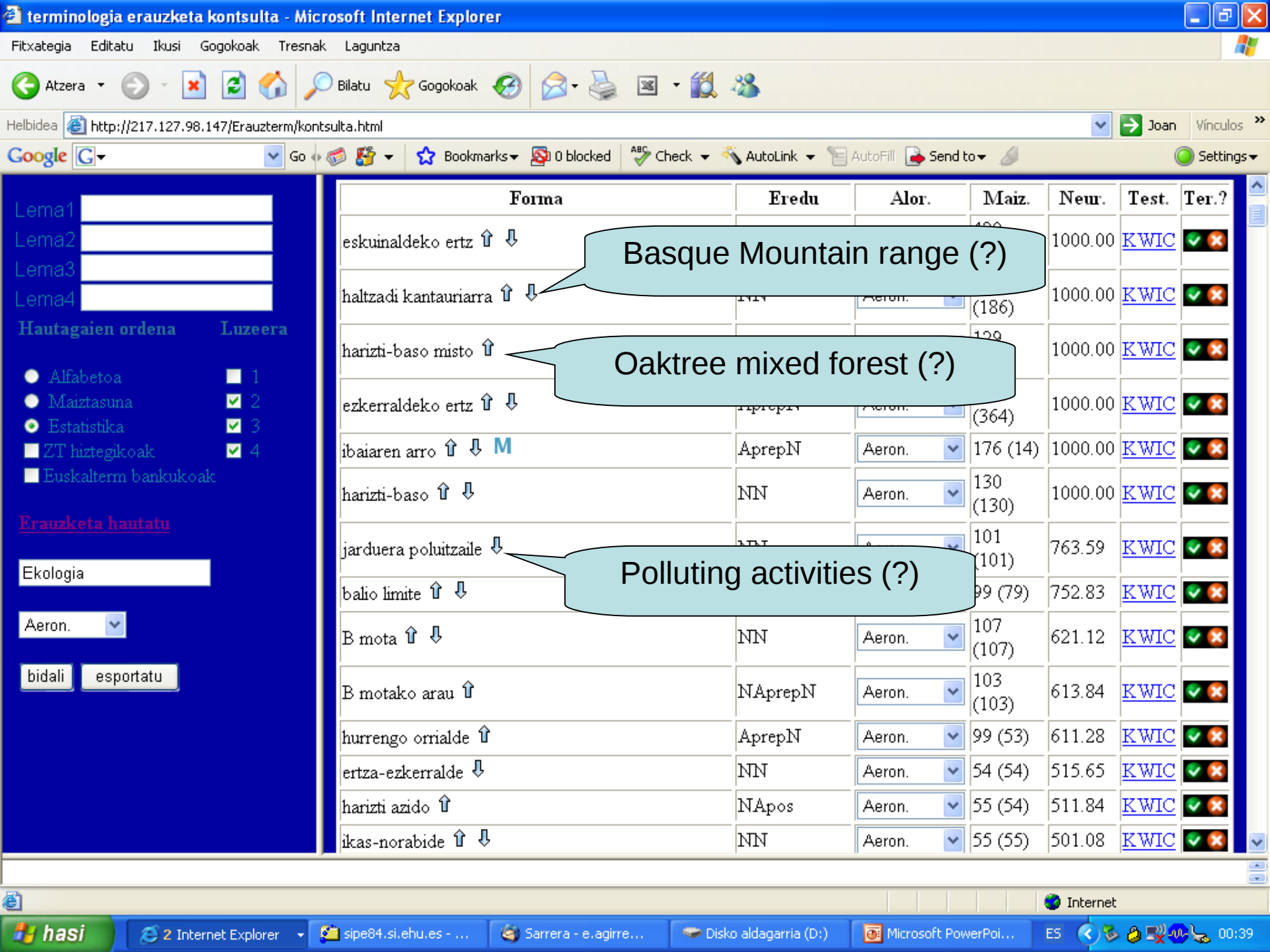
```
<!-- every -->
<quantifiers>
  <quantifier qid=”q1” span=”t5”/>
</quantifiers>

<!-- 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


File View Help
C:\Irion\Twentyone 3.0\Index 3.1\Projects\Kyoto\Kyoto_3\1\classes_en.xml
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...
100%


terminologia erauzketa kontsulta - Microsoft Internet Explorer
Fitxategia Editatu Ikusi Gogokoak Tresnak Laguntza
Atzera Bilatu Gogokoak
Helbidea http://217.127.98.147/Erauzterm/kontsulta.html
Joan Vínculos
Google Go Bookmarks 0 blocked Check AutoLink AutoFill Send to Settings
Lema1
Lema2
Lema3
Lema4
Hautagaien ordena
Luzeera
Alfabetoa
Maiztasuna
Estatistika
ZT hiztegikoak
Euskalterm bankukoak
1
2
3
4
Erauzketa hautatu
Ekologia
Aeron.
bidali
esportatu
Forma
Eredu
Alor.
Maiz.
Neur.
Test.
Ter.?
eskuinaldeko ertz
haltzadi kantauriarra
harizti-baso misto
ezkerraldeko ertz
ibaiaren arro
harizti-baso
jarduera poluitzaile
balio limite
B mota
B motako arau
hurrengo orrialde
ertza-ezkerralde
harizti azido
ikas-norabide
AprepN
NN
NAprepN
NApos
Aeron.
176 (14)
130 (130)
101 (101)
99 (79)
107 (107)
103 (103)
99 (53)
54 (54)
55 (54)
55 (55)
1000.00
763.59
752.83
621.12
613.84
611.28
515.65
511.84
501.08
KWIC
Basque Mountain range (?)
Oaktree mixed forest (?)
Polluting activities (?)
Internet
hasi
2 Internet Explorer
sipe84.si.ehu.es - ...
Sarrera - e.agirre...
Disko aldagarria (D:)
Microsoft PowerPoi...
ES
00:39

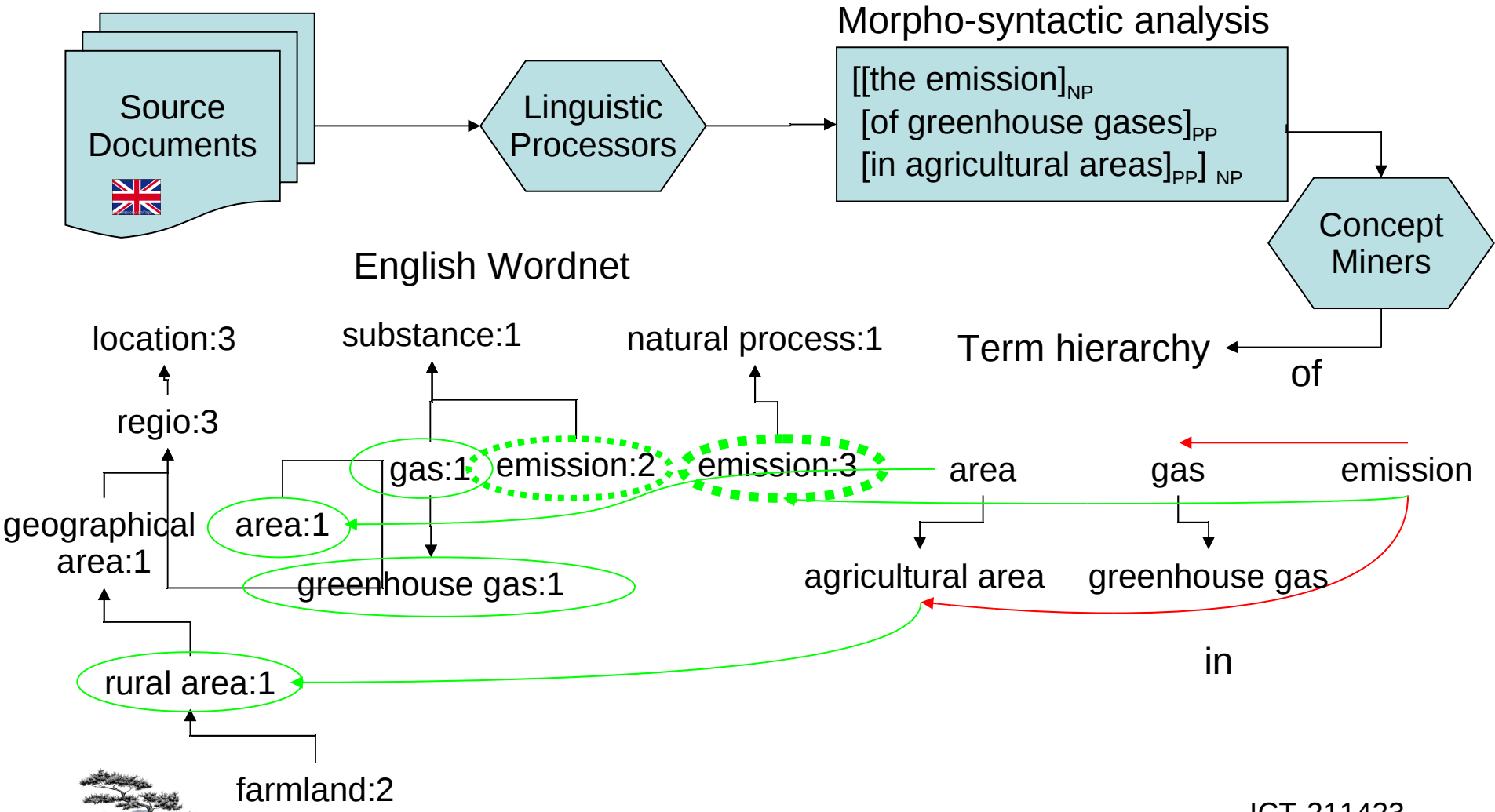
Infomap BNC + SSI-Dijkstra

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



Kybots, knowledge yielding robots

- What kybots do?
- Mining module architecture
- Kybot profiles
  - Current capabilities
- Running kybots
  - XQuery
  - Performance
- Building Kybots
  - Mining by example
  - Machine Learning / Active Learning
- Next steps

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**)
  - Define morpho-syntactic and semantic patterns in text
  - Extract events from text
  - Collect events and extract facts

- For all languages!
- KAF (Kyoto Annotation Format) is the input of both:
  - Tybot: term extraction
  - Kybot: fact extraction

Linguistic Processors

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

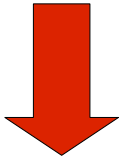
- MW detection: **VUA**
- Word Sense Disambiguation module (UKB): **EHU**
- NE Tagger: **Irion**
- OntoTagger: **CNR-ILC**, **EHU**

Linguistic Processors

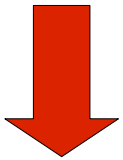
- KAF XML files include sections for:
  - Word forms
  - Terms / Items
  - Chunks: grouping of sequences of terms
  - Dependencies: syntactic relations between terms
  - WSD: WN senses of the term
  - Ontological references of the term:
    - Base Concepts
    - Explicit ontology
  - Events
  - *Quantifiers, Time expressions, General Relations*
  - ...

Fact Mining: Kybots

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



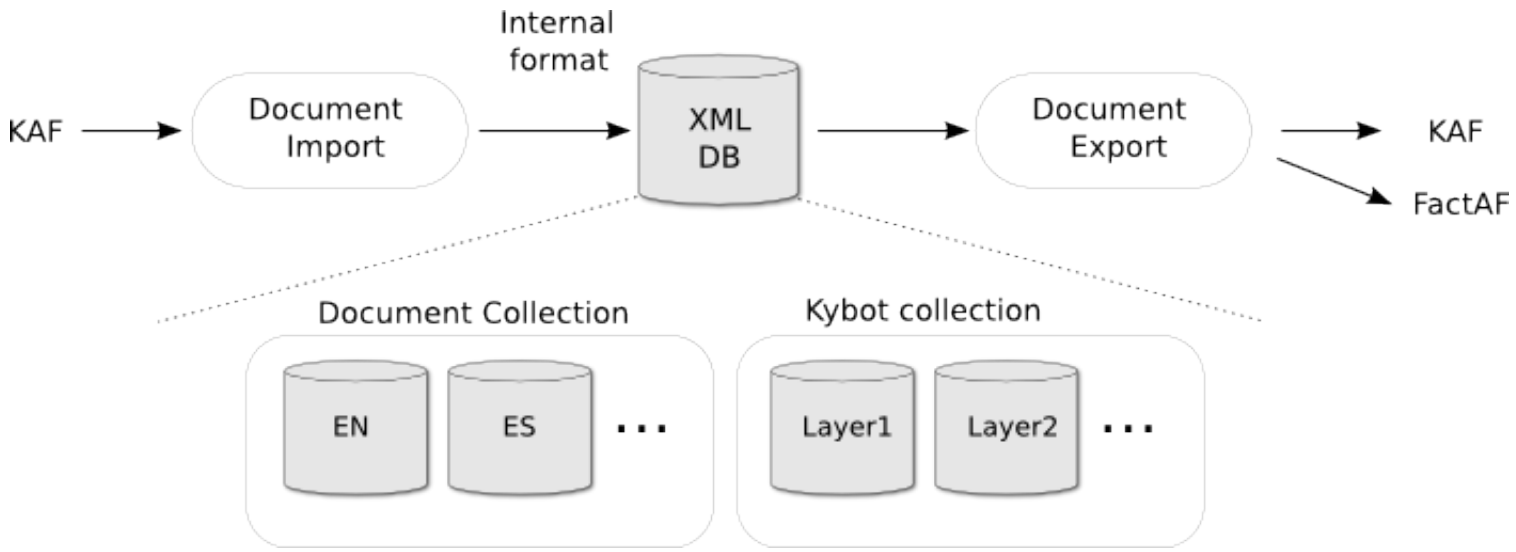
+ Linguistic Processing: POS, chunks, dependencies, ...
+ Semantic Processing: WSD (=>WN => ontology)



+ Kybot profiles: morphosyntactic + semantic patterns
+ Mining Module: Events / Facts

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

Mining Module Architecture



- Central XML DB stores
  - Documents (in all languages)
  - Kybots (organized in libraries)
- Kybots are executed using Xqueries on the XML DB

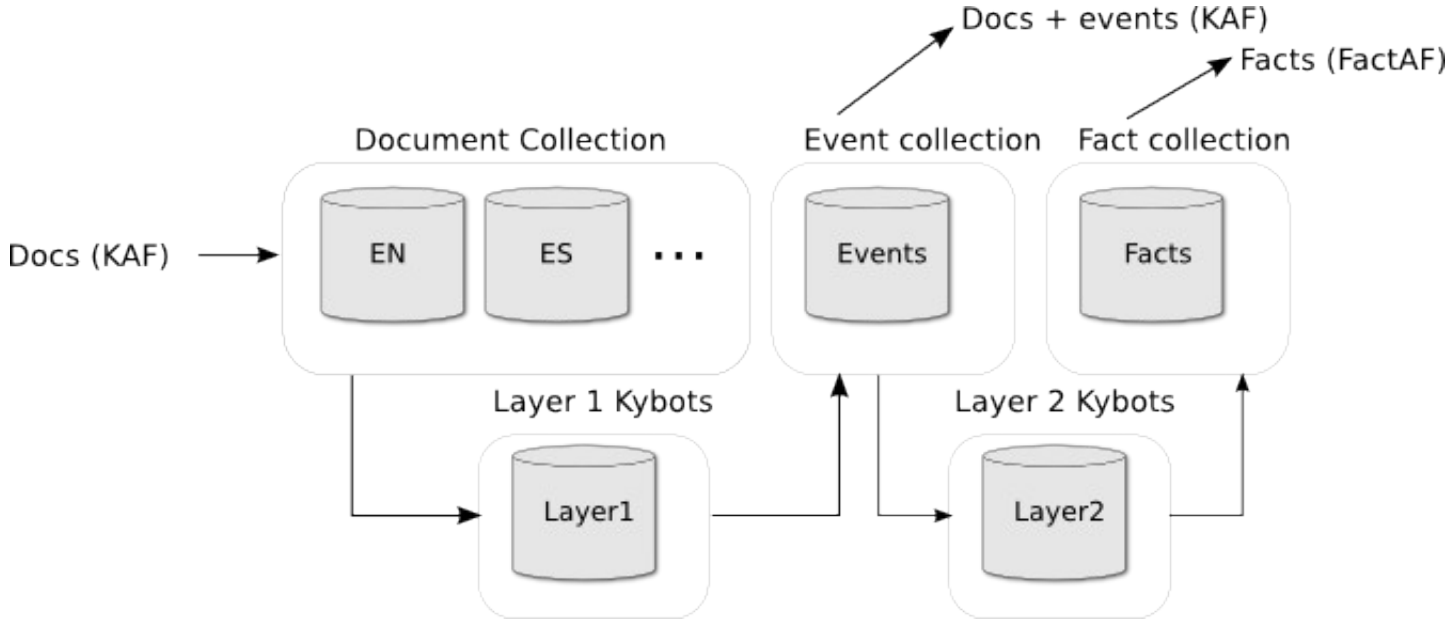
Mining Module capabilities

- Load KAF documents
  - Converts KAF to internal representation
    - Explicit boundaries: sentence, paragraphs, etc.
    - Indexing
- Exporting to KAF
- Application of Kybots
- Listing content
- ...

Kybot application

- User uploads a Kybot profile to the collection
- User applies a Kybot (Kybot-pipeline) to Docs
  - Or a subset of docs (ex. only a language)
- Some Kybots add information to existing Docs
  - Events (layer 1)
- Some Kybots create new facts
  - FactAF (layer 2)
- Also, keep track of which kybot created which fact

Layered Kybots



Layer 1 Kybots:

Input: KAF (+ MW, WSD, NE, Ontological Information)

Output: events (and roles)

Layer 2 Kybots:

Input: events (from different docs, languages)

Output: facts

Kybot profiles

- Use XML syntax to define the kybots
- Self descriptive (for manual Kybot creation)
- Powerful expressions
  - terms:
    - POS
    - Lemma
    - Senses, Base Concepts
    - Ontological references
  - suffix/prefix expressions
  - conjunction, disjunction, optionality
  - Negation
- Efficient
  - Able to manage thousands of KAF documents

Fact Mining: Kybot profiles

- Kybot profiles consist of:
  - Expression Rules
    - Morpho-syntactic conditions on the LPs outcomes
    - Flexible enough for dealing with all KAF outputs
  - Semantic conditions:
    - WordNets + Ontologies
    - Inferencing on WN / ontology !
  - Output Template
    - Event / Fact descriptions

Fact Mining: Kybot profiles

- For each analysed sentence :
  - **IF**
    - Expression Rules match **and**
    - Semantic Conditions hold
  - **THEN**
    - generate the Output Template

- How to make efficient inferencing on WN / ontology?
  - ... while processing very large volumes of KAF

  - WN => Nominal and Verbal Base Concepts !
  - Ontology => Explicit Ontology !

Kybot profiles

```
<?xml version="1.0" encoding="utf-8"?>

<Kybot id="Generate_Pollution">

<variables>
    <var name="X" type="term" pos="N"/>
    <var name="Y" type="term" lemma="release | produce | generate | ! create"/>
    <var name="Z" type="term" lemma="*pollution | pollutant | contaminant"/>
</variables>

<relations>
    <root span="X"/>
    <rel span="Y" pivot="X" direction="following"/>
    <rel span="Z" pivot="Y" direction="following"/>
</relations>

<events>
    <event target="$Y/@tid" lemma="$Y/@lemma" pos="$Y/@pos"/>
    <role target="$X/@tid" rtype="source" lemma="$X/@lemma" pos="$X/@pos"/>
    <role target="$Z/@tid" rtype="patient" lemma="$Z/@lemma" pos="$Z/@pos"/>
</events>

</Kybot>
```

Kybot profiles

```
<?xml version="1.0" encoding="utf-8"?>

<Kybot id="Generate_Pollution">

<variables>                                                    Variables
    <var name="X" type="term" pos="N"/>
    <var name="Y" type="term" lemma="release | produce | generate | ! create"/>
    <var name="Z" type="term" lemma="*pollution | pollutant | contaminant"/>
</variables>

<relations>
    <root span="X"/>
    <rel span="Y" pivot="X" direction="following"/>
    <rel span="Z" pivot="Y" direction="following"/>
</relations>

<events>
    <event target="$Y/@tid" lemma="$Y/@lemma" pos="$Y/@pos"/>
    <role target="$X/@tid" rtype="source" lemma="$X/@lemma" pos="$X/@pos"/>
    <role target="$Z/@tid" rtype="patient" lemma="$Z/@lemma" pos="$Z/@pos"/>
</events>

</Kybot>
```

Kybot profiles

```
<?xml version="1.0" encoding="utf-8"?>

<Kybot id="Generate_Pollution">

<variables>
    <var name="X" type="term" pos="N"/>
    <var name="Y" type="term" lemma="release | produce | generate | ! create"/>
    <var name="Z" type="term" lemma="*pollution | pollutant | contaminant"/>
</variables>

<relations>
    <root span="X"/>
    <rel span="Y" pivot="X" direction="following"/>
    <rel span="Z" pivot="Y" direction="following"/>
</relations>

<events>
    <event target="$Y/@tid" lemma="$Y/@lemma" pos="$Y/@pos"/>
    <role target="$X/@tid" rtype="source" lemma="$X/@lemma" pos="$X/@pos"/>
    <role target="$Z/@tid" rtype="patient" lemma="$Z/@lemma" pos="$Z/@pos"/>
</events>

</Kybot>
```

Relations

Kybot profiles

```
<?xml version="1.0" encoding="utf-8"?>

<Kybot id="Generate_Pollution">

<variables>
    <var name="X" type="term" pos="N"/>
    <var name="Y" type="term" lemma="release | produce | generate | ! create"/>
    <var name="Z" type="term" lemma="*pollution | pollutant | contaminant"/>
</variables>

<relations>
    <root span="X"/>
    <rel span="Y" pivot="X" direction="following"/>
    <rel span="Z" pivot="Y" direction="following"/>
</relations>

<events>
    <event target="$Y/@tid" lemma="$Y/@lemma" pos="$Y/@pos"/>
    <role target="$X/@tid" rtype="source" lemma="$X/@lemma" pos="$X/@pos"/>
    <role target="$Z/@tid" rtype="patient" lemma="$Z/@lemma" pos="$Z/@pos"/>
</events>
```

Output Template

```
</Kybot>
```

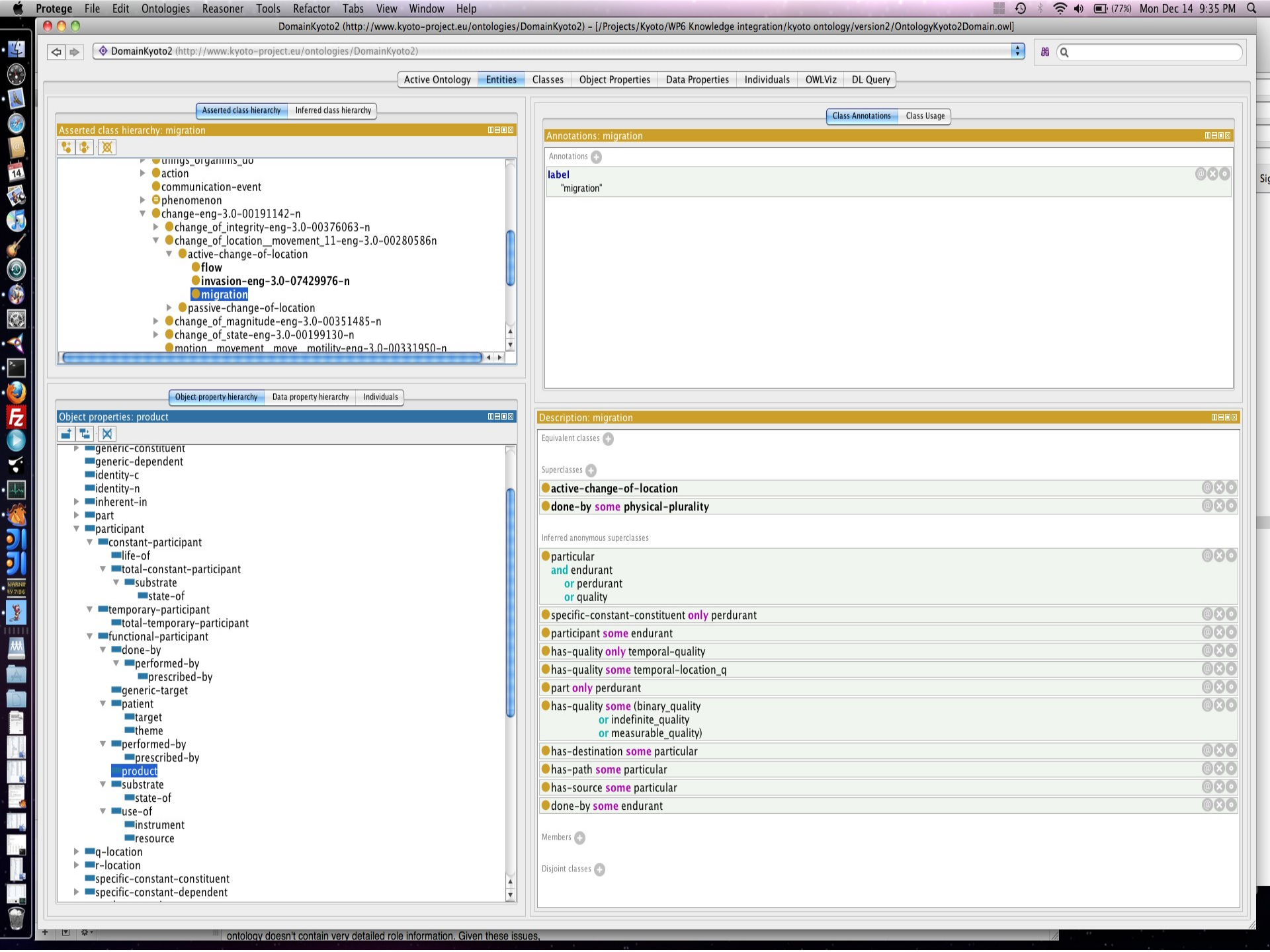
Kybot profiles: Output

```
<?xml version="1.0"?>
<kybotOut>
  <doc shortname="1534.mw.wsd.ne.onto.kaf">
    <event target="t886" lemma="generate" pos="V" eid="e1"/>
    <role target="t884" rtype="source" lemma="watershed" .../>
    <role target="t892" rtype="patient" lemma="pollution" .../>
  </doc>
  <doc shortname="17795.mw.wsd.ne.onto.kaf">
    <event target="t9690" lemma="release" pos="V" eid="e1"/>
    <role target="t9691" rtype="patient" lemma="pollutant" .../>
    <role target="t9678" rtype="source" lemma="fuel" .../>
    <role target="t9680" rtype="source" lemma="heating" .../>
    <role target="t9681" rtype="source" lemma="machinery" .../>
    <role target="t9683" rtype="source" lemma="equipment" .../>
    <role target="t9686" rtype="source" lemma="household" .../>
    <role target="t9688" rtype="source" lemma="business" .../>
  </doc>
</kybotOut>
```

Kybot profiles: Ontological references

```
<?xml version="1.0" encoding="utf-8"?>
<!-- N produces changes of pollution level -->

<Kybot id="Changes_Pollution">
<variables>
    <var name="A" type="term" pos="N"/>
    <var name="B" type="term"
         reference="Kyoto#measure__quantity__amount-eng-3.0-00033615-n"
         reftype="SubClassOf"/>
    <var name="C" type="term" pos="P"/>
    <var name="D" type="term"
         reference="Kyoto#contamination__pollution-eng-3.0-00276987-n"
         reftype="SubClassOf"/>
</variables>
<relations>
    <root span="D"/>
    <rel span="C" pivot="D" direction="preceding"/>
    <rel span="B" pivot="C" direction="preceding"/>
    <rel span="A" pivot="B" direction="preceding"/>
</relations>
<events>
    <event target="$B/@tid" lemma="$B/@lemma" pos="$B/@pos"/>
    <role target="$A/@tid" rtype="agent" lemma="$A/@lemma"
          pos="$A/@pos"/>
    <role target="$D/@tid" rtype="patient" lemma="$D/@lemma"
          pos="$D/@pos"/>
</events>
</Kybot>
```


Protege File Edit Ontologies Reasoner Tools Refactor Tabs View Window Help
Mon Dec 14 9:35 PM
DomainKyoto2 (http://www.kyoto-project.eu/ontologies/DomainKyoto2) – [/Projects/Kyoto/WP6 Knowledge integration/kyoto ontology/version2/OntologyKyoto2Domain.owl]
DomainKyoto2 (http://www.kyoto-project.eu/ontologies/DomainKyoto2)
Active Ontology
Entities
Classes
Object Properties
Data Properties
Individuals
OWLViz
DL Query
Asserted class hierarchy
Inferred class hierarchy
Asserted class hierarchy: migration
things_organisms_do
action
communication-event
phenomenon
change-eng-3.0-00191142-n
change_of_integrity-eng-3.0-00376063-n
change_of_location__movement_11-eng-3.0-00280586n
active-change-of-location
flow
invasion-eng-3.0-07429976-n
migration
passive-change-of-location
change_of_magnitude-eng-3.0-00351485-n
change_of_state-eng-3.0-00199130-n
motion__movement__move__motility-eng-3.0-00331950-n
Class Annotations
Class Usage
Annotations: migration
Annotations
label
"migration"
Object property hierarchy
Data property hierarchy
Individuals
Object properties: product
generic-constituent
generic-dependent
identity-c
identity-n
inherent-in
part
participant
constant-participant
life-of
total-constant-participant
substrate
state-of
temporary-participant
total-temporary-participant
functional-participant
done-by
performed-by
prescribed-by
generic-target
patient
target
theme
performed-by
prescribed-by
product
substrate
state-of
use-of
instrument
resource
q-location
r-location
specific-constant-constituent
specific-constant-dependent
Description: migration
Equivalent classes
Superclasses
active-change-of-location
done-by some physical-plurality
Inferred anonymous superclasses
particular
and endurant
or perdurant
or quality
specific-constant-constituent only perdurant
participant some endurant
has-quality only temporal-quality
has-quality some temporal-location_q
part only perdurant
has-quality some (binary_quality
or indefinite_quality
or measurable_quality)
has-destination some particular
has-path some particular
has-source some particular
done-by some endurant
Members
Disjoint classes
ontology doesn't contain very detailed role information. Given these issues,

Explicit Ontology

- Explicit knowledge:
  - Kyoto#migration SubClassOf Kyoto#active-change-of-location
  - Kyoto#migration Kyoto#done-by Collections.owl#physical-plurality

- Implicit knowledge:
  - Kyoto#migration SubClassOf Kyoto#change_of_location__movement_11-eng-3.0-00280586n inherited
  - Kyoto#migration SubClassOf Kyoto#change-eng-3.0-00191142-n inherited
  - Kyoto#migration SubClassOf DOLCE-Lite.owl#accomplishment inherited
  - Kyoto#migration SubClassOf DOLCE-Lite.owl#event inherited
  - Kyoto#migration SubClassOf DOLCE-Lite.owl#perdurant inherited
  - Kyoto#migration SubClassOf DOLCE-Lite.owl#spatio-temporal-particular inherited
  - Kyoto#migration SubClassOf DOLCE-Lite.owl#particular inherited
  - Kyoto#migration Kyoto#has-path DOLCE-Lite.owl#particular inherited
  - Kyoto#migration Kyoto#has-destination DOLCE-Lite.owl#particular inherited
  - Kyoto#migration Kyoto#has-source DOLCE-Lite.owl#particular inherited
  - Kyoto#migration DOLCE-Lite.owl#has-quality DOLCE-Lite.owl#temporal-location_q inherited
  - Kyoto#migration DOLCE-Lite.owl#specific-constant-constituent DOLCE-Lite.owl#perdurant inherited
  - Kyoto#migration DOLCE-Lite.owl#participant DOLCE-Lite.owl#endurant inherited
  - Kyoto#migration DOLCE-Lite.owl#part DOLCE-Lite.owl#perdurant inherited
  - Kyoto#migration DOLCE-Lite.owl#has-quality DOLCE-Lite.owl#temporal-quality inherited

Kybot profiles: Output

```
<?xml version="1.0"?>
<kybotOut>
  <doc shortname="11767.mw.wsd.ne.onto.kaf">
    <event target="t3494" lemma="be" pos="V" eid="e1"/>
    <role target="t3493" rtype="agent" lemma="pollution" .../>
    <role target="t3504" rtype="patient" lemma="industrial_facility" .../>
    <event target="t3687" lemma="change" pos="N" eid=”e2”/>
    <role target="t3683" rtype="agent" lemma="precipitation" .../>
    <role target="t3690mw" rtype="patient" lemma="pollution level" .../>
    <event target="t3737" lemma="be" pos="V" eid="e3"/>
    <role target="t3736" rtype="agent" lemma="pipe" .../>
    <role target="t3742mw" rtype="patient" lemma="pollution level" .../>
    <event target="t5833" lemma="change" pos="V" eid="e4"/>
    <role target="t5826" rtype="agent" lemma="alga" .../>
    <role target="t5836mw" rtype="patient" lemma="pollution level" />
    <event target="t7378" lemma="be" pos="V" eid="e5"/>
    <role target="t7377" rtype="agent" lemma="there" />
    <role target="t7383mw" rtype="patient" lemma="polluted stream" .../>
  </doc>
</kybotOut>
```

Kybot profiles: Performance

- Benckmark database
  - 3 documents
  - 26,137 word forms
  - 96Mb KAF documents, 741Mb dbxml index

- Estuary database
  - 4.624 documents
  - 3,091,181 word forms
  - 8.2Gb KAF documents, 45Gb dbxml index

Next steps

- Selecting the most appropriate senses
- Improve KAF representation for explicit ontology
  - Ontology concepts are coarser than senses
- Chunk level queries
  - Search for a term and then a chunk whose head is ...
  - Inter-chunk searches
    - Search for a term and then, in the same chunk, another one which ...
- Layer-2 Kybots
  - Amalgamate events from several documents and languages
- Generic Kybots
- Creating Kybots
  - Mining by example
  - Machine learning / Active Learning

Generic Kybots: Kybots and Ontology

- Ontology
  - Events, roles, Fillers
  - e.g. BirdMigration (role: agent; filler: Bird)
- Linguistic realizations:
  - *migration of birds*
  - *bird migration*
  - *birds migrate*
  - ...
- OntoTagger:
  - “migration” --> BirdMigration event (role: agent; filler: Bird)
  - “migrate” --> BirdMigration event (role: agent; filler: Bird)
  - “robin” --> Bird

Generic Kybot: Rules

- *migration of birds*
- N1 **of** N2 --> O1 event (role: agent; filler: O2) **IF**
  N1 is event O1 **AND**
  N2 is concept O2 **AND**
  O1 has role agent **AND**
  O2 is (*subsumed by*) filler of agent of O1

- *robin migrate*
- N V --> O1 event (role: agent; filler: O2) **IF**
  V is event O1 **AND**
  N is concept O2 **AND**
  O1 has role agent **AND**
  O2 is (*subsumed by*) filler of agent of O1

Building Kybots: Mining by example

- Kybots perform a complex **Information Extraction** (IE) task requiring expertise on:
  - linguistic engineering
  - knowledge engineering ...
- but ...
  - all this complexity could 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, inferencing, etc.

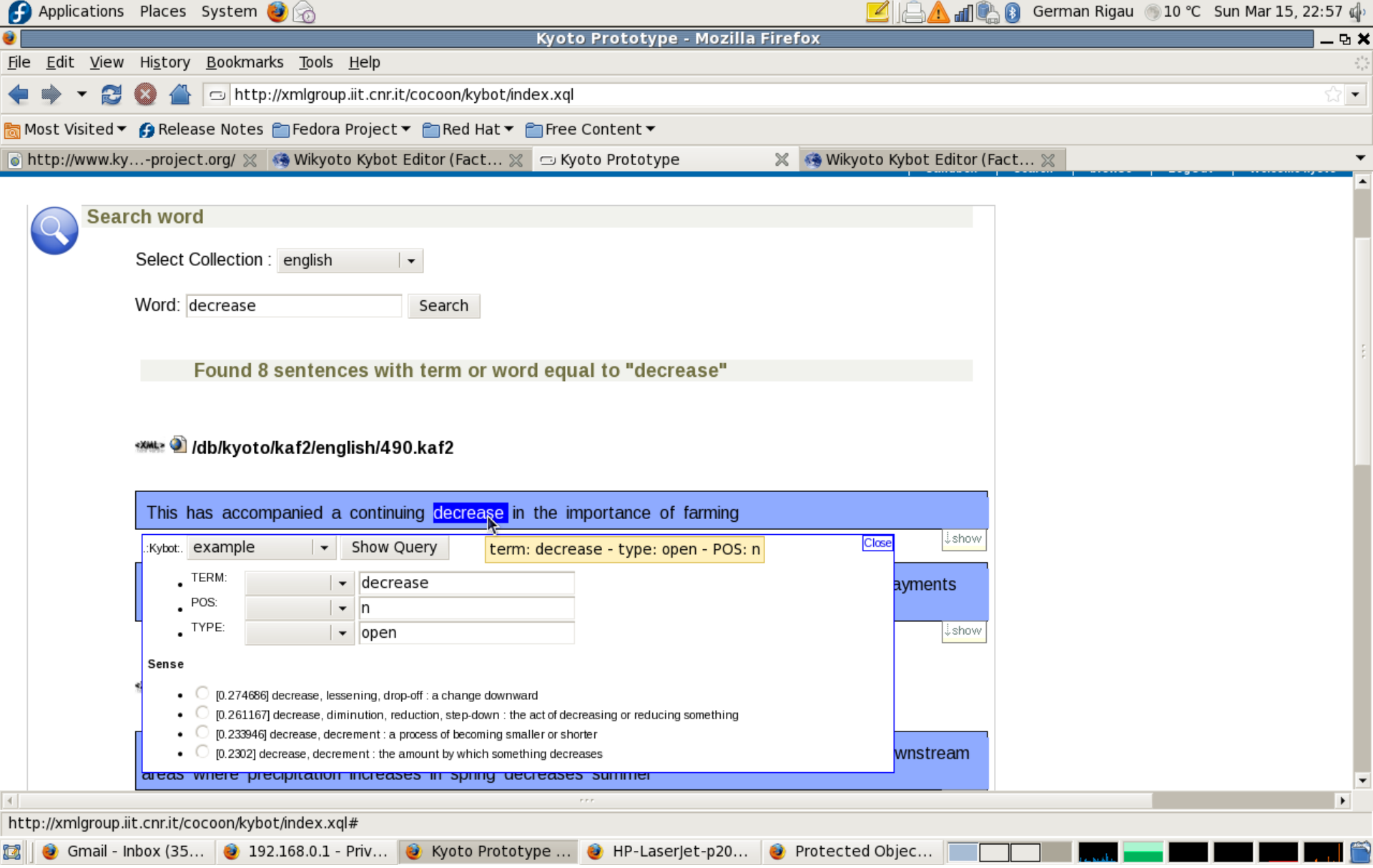
Building Kybots: Mining by example

- 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

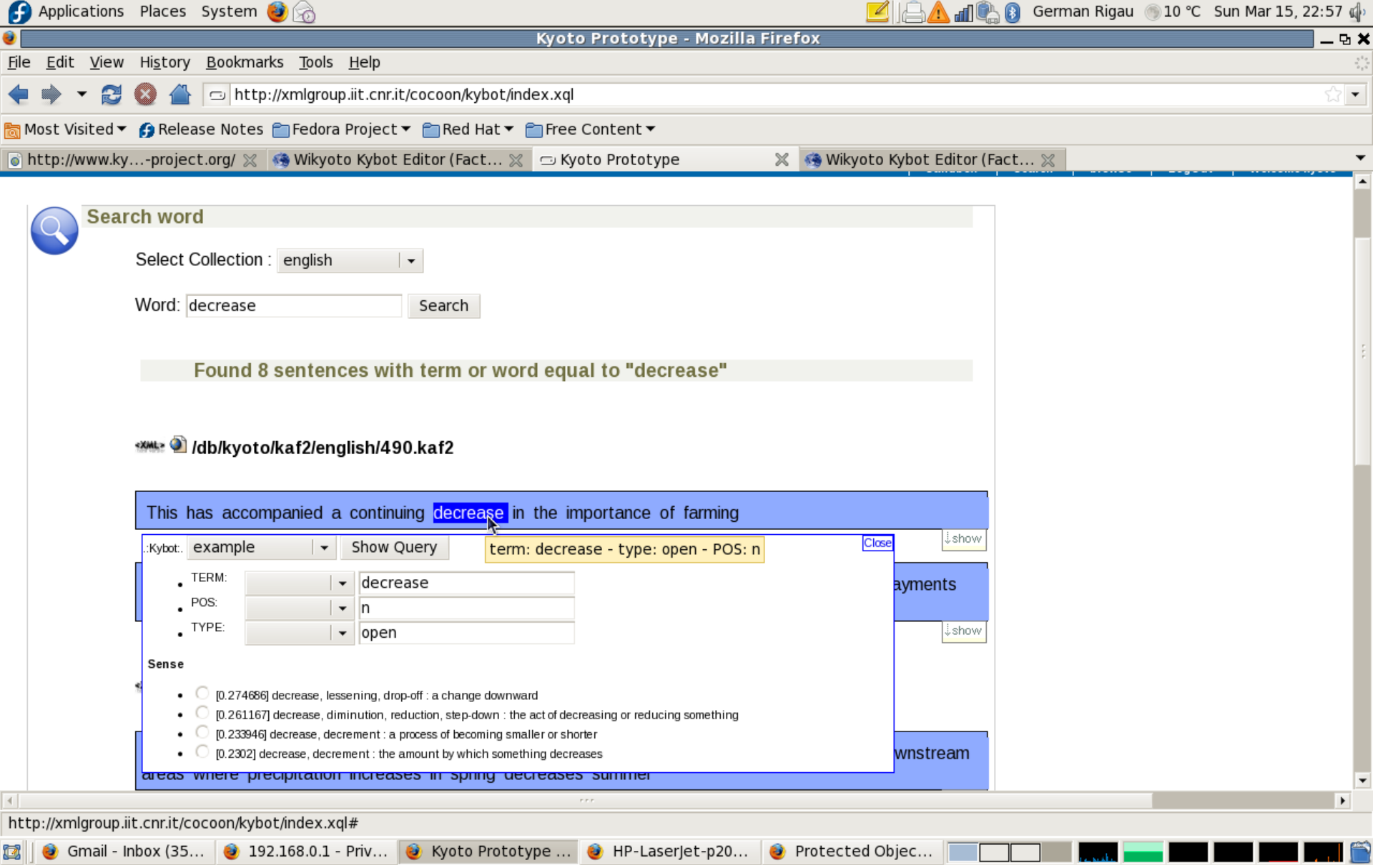
Building Kybots: Mining by example

- "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
  - Wordnet information: Base Concepts, ...
  - Ontological information, ...
  - ...



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



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.

search...
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

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 Attributes 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 Attributes 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
- - - - - - -
Contact Us
Template by Mambo Hut
Login Form
Username
Password
Remember me
Login
Password Reminder
No account yet?
Create one

Frame Elements and Their Syntactic Realizations

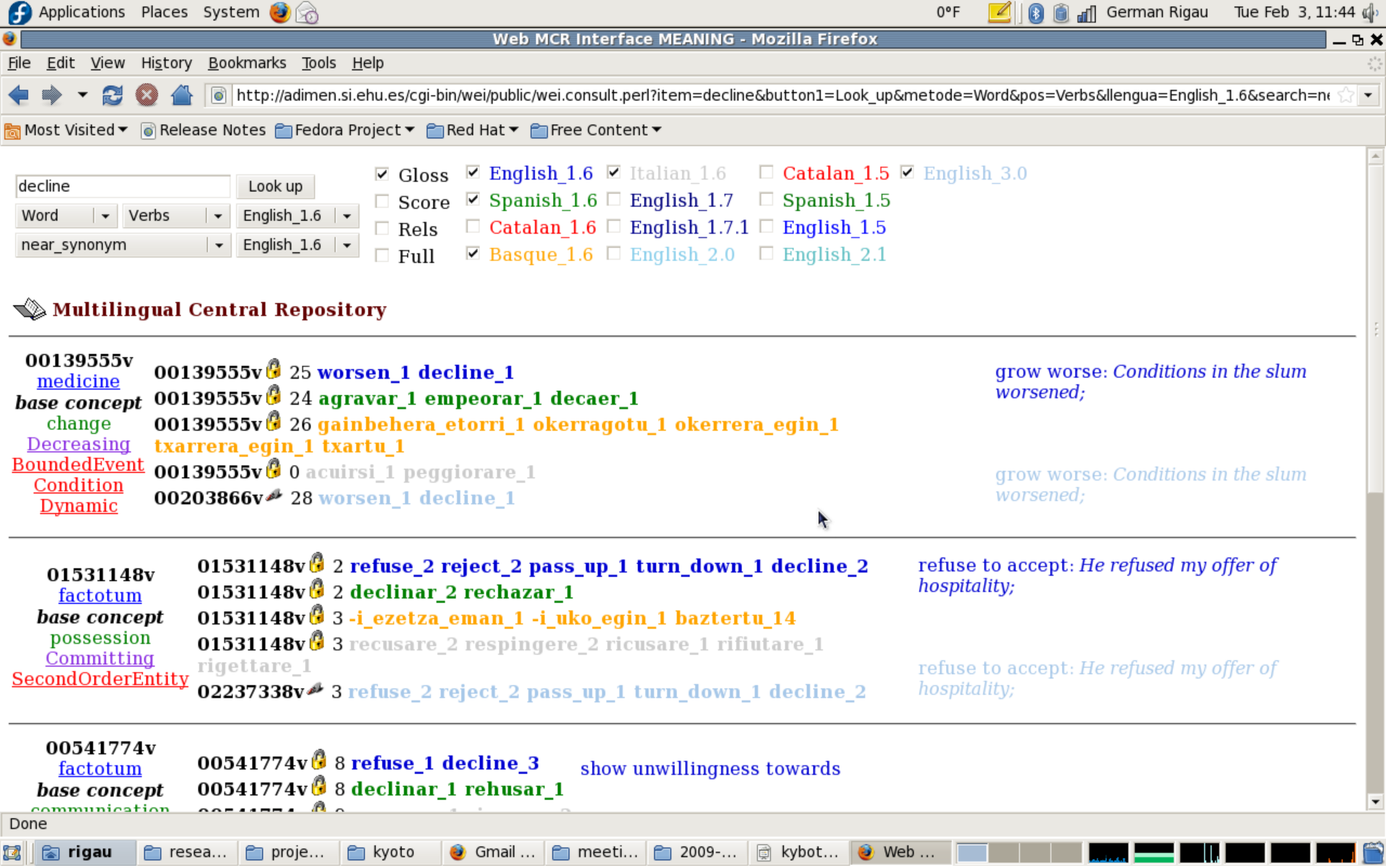
The Frame elements for this word sense are (with realizations):

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		



http://adimen.si.ehu.es/web/MCR

Building Kybots: Mining by example

- 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, base concepts, ontologies ...
    - The class of verb of change followed by preposition followed by...
    - ...

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
  - Eficiency vs. expressivity
  - Internal KAF representation for efficiency / indexing
  - Combination of kybots
  - ...

KYOTO (ICT-211423) Intelligent Content and Semantics
Knowledge **Y**ielding **O**ntologies for **T**ransition-Based **O**rganization
http://www.kyoto-project.eu/

Kybots, knowledge yielding robots

German Rigau
IXA group, UPV/EHU

First Review Meeting
March 17, 2009, Luxembourg