KYOT0 (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
KYOTO Overview

- **Title:** Knowledge Yielding Ontologies for Transition-Based Organization
- **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

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

Environmental organizations

1. ECNC
2. WWF
3. Wikyoto

Capture text:
"Sudden increase of CO2 emissions in 2008 in Europe"

Tybot: term yielding robot

CO2 emission

Kybot: knowledge yielding robot

Index facts:
- Process: Emission
- Involves: CO2
- Property: Increase, sudden
- When: 2008
- Where: Europe

Wordnets

Ontology

Top

Abstract Physical

Process Substance

H20 CO2

Middle

Pollution Emission Greenhouse Gas

Domain

H20 CO2

 Semantic 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
Term Editor (Wikyoto)

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

Domain Wordnet
K-LMF

Domain Ontology
OWL-DL

Concept User

Extracted Terms
Generic K-TMF

Fact User

Kybot Editor

Kybot Profiles

Kybot Server
(Fact Extraction)

Document Base
KAF

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

Expression rules:
- N+N
- N+prep+N
- Nsubj+V
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>
“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>
</terms>
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>  
</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
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>eskuinaldeko etzt</td>
<td>NN</td>
<td>NN</td>
<td>1000.0</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>haltzadi kantauriarrira</td>
<td>NN</td>
<td>NN</td>
<td>1000.0</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>harizti-baso misto</td>
<td>NN</td>
<td>NN</td>
<td>1000.0</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ezkerraldeko etzt</td>
<td>NN</td>
<td>NN</td>
<td>1000.0</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ibairen arro</td>
<td>AprepN</td>
<td>Aeron.</td>
<td>176 (14)</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>harizti-baso</td>
<td>NN</td>
<td>NN</td>
<td>1000.0</td>
<td>KWIC</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>jarduera poluitzaile</td>
<td>NN</td>
<td>NN</td>
<td>101</td>
<td>763.59</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>balio limite</td>
<td>NN</td>
<td>NN</td>
<td>99 (79)</td>
<td>752.83</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>B mota</td>
<td>NN</td>
<td>NN</td>
<td>107 (107)</td>
<td>621.12</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>B motako arau</td>
<td>NAprepN</td>
<td>Aeron.</td>
<td>103 (103)</td>
<td>613.84</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>hurrengo orrialde</td>
<td>AprepN</td>
<td>Aeron.</td>
<td>99 (53)</td>
<td>611.28</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>etza-ekzerralde</td>
<td>NN</td>
<td>NN</td>
<td>54</td>
<td>515.65</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>harizti azido</td>
<td>NApos</td>
<td>Aeron.</td>
<td>55 (54)</td>
<td>511.84</td>
<td>KWIC</td>
<td>✓</td>
</tr>
<tr>
<td>ikas-norabide</td>
<td>NN</td>
<td>NN</td>
<td>55 (55)</td>
<td>501.08</td>
<td>KWIC</td>
<td>✓</td>
</tr>
</tbody>
</table>

- **Basque Mountain range**
- **Oaktree mixed forest**
- **Polluting activities**
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

Source Documents

Linguistic Processors

Morpho-syntactic analysis

[[the emission]_{NP}
[of greenhouse gases]_{PP}
[in agricultural areas]_{PP}]_{NP}

Concept Miners

English Wordnet

location:3
regio:3
geographical area:1
rural area:1

substance:1
natural process:1

emission:2
emission:3

gas:1

emission
greenhouse gas:1
agricultural area

Term hierarchy

area
gas
emission
in
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*
  - ...

ICT-211423
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)

KAF

+ 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
- ...

ICT-211423
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>
    <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>
<?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>
<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: Output

<?xml version="1.0"?>
<kybot Out >
  <doc short name="1534. mw. wsd. ne. ont o. 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 short name="17795. mw. wsd. ne. ont o. 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>
</kybot Out>
Kybot profiles: Ontological references

```xml
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE Kybot [ ]>

<!-- 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#measur e_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="preceeding"/>
    <rel span="B" pivot="C" direction="preceeding"/>
    <rel span="A" pivot="B" direction="preceeding"/>
  </relations>
  <events>
    <event target="$B/@id" lemma="$B/@lemma" pos="$B/@pos"/>
    <role target="$A/@id" rtype="agent" lemma="$A/@lemma" pos="$A/@pos"/>
    <role target="$D/@id" rtype="patient" lemma="$D/@lemma" pos="$D/@pos"/>
  </events>
</Kybot>
```
Ontology

- events
- roles
- Fillers
  - e.g. BirdMigration (role: agent; filler: Bird)

Linguistic realizations:
- migration of birds
- bird migration
- birds migrate
- ...
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="algae"/>
  <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

- Benchmark 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, ...
  - ...

ICT-211423
This has accompanied a continuing decrease in the importance of farming.
This has accompanied a continuing decrease in the importance of farming.
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.
**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 NPs 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 not 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.
2. (1-second FE layer)Accidents **INCREASED** 20% to 345.
3. **Attacks on civilians** **DECREASED** over the last 4 months.
4. DNI
## Frame Elements and Their Syntactic Realizations

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

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

## Valence Patterns:

These frame elements occur in the following syntactic patterns:

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 TOTAL</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multilingual Central Repository

00139555v 00139555v 25 worsen_1 decline_1
grow worse: Conditions in the slum
worsened;

base concept
medicine
change
BoundedEvent
Condition
Dynamic

00139555v 00139555v 24 agravar_1 empeorar_1 decaer_1
00139555v 00139555v 26 gainbehera_etorri_1 okerragotu_1 okerrera_egin_1
txarrera_egin_1 txartu_1
00139555v 00139555v 0 acurri_1 peggiorare_1
grow worse: Conditions in the slum
worsened;

01531148v 01531148v 2 refuse_2 reject_2 pass_up_1 turn_down_1 decline_2
refuse to accept: He refused my offer of
hospitality;

factotum
base concept
possesion
Committing
SecondOrderEntity

01531148v 01531148v 2 declarar_2 rechazar_1
01531148v 01531148v 3 -i_ozetza_eman_1 -i_uko_egin_1 baztertu_14
01531148v 01531148v 3 recusare_2 respingere_2 ricusare_1 rifiutare_1
rigettare_1
02237338v 02237338v 3 refuse_2 reject_2 pass_up_1 turn_down_1 decline_2
refuse to accept: He refused my offer of
hospitality;

00541774v 00541774v 8 refuse_1 decline_3
show unwillingness towards

factotum
base concept
communication

00541774v 00541774v 8 declarar_1 rehusar_1

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

ICT-211423
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...
    - ...

ICT-211423
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
  - ...

ICT-211423
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