KYOTO (ICT-211423)
Yielding **O**ntologies for **T**ransition-Based **O**rganization
FP7: Intelligent Content and Semantics

http://www.kyoto-project.eu/

Piek Vossen

KYOTO (ICT-211423) Overview

- **Title: 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
  - Enables knowledge transition and information search 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.

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)

KYOTO (ICT-211423) Overview

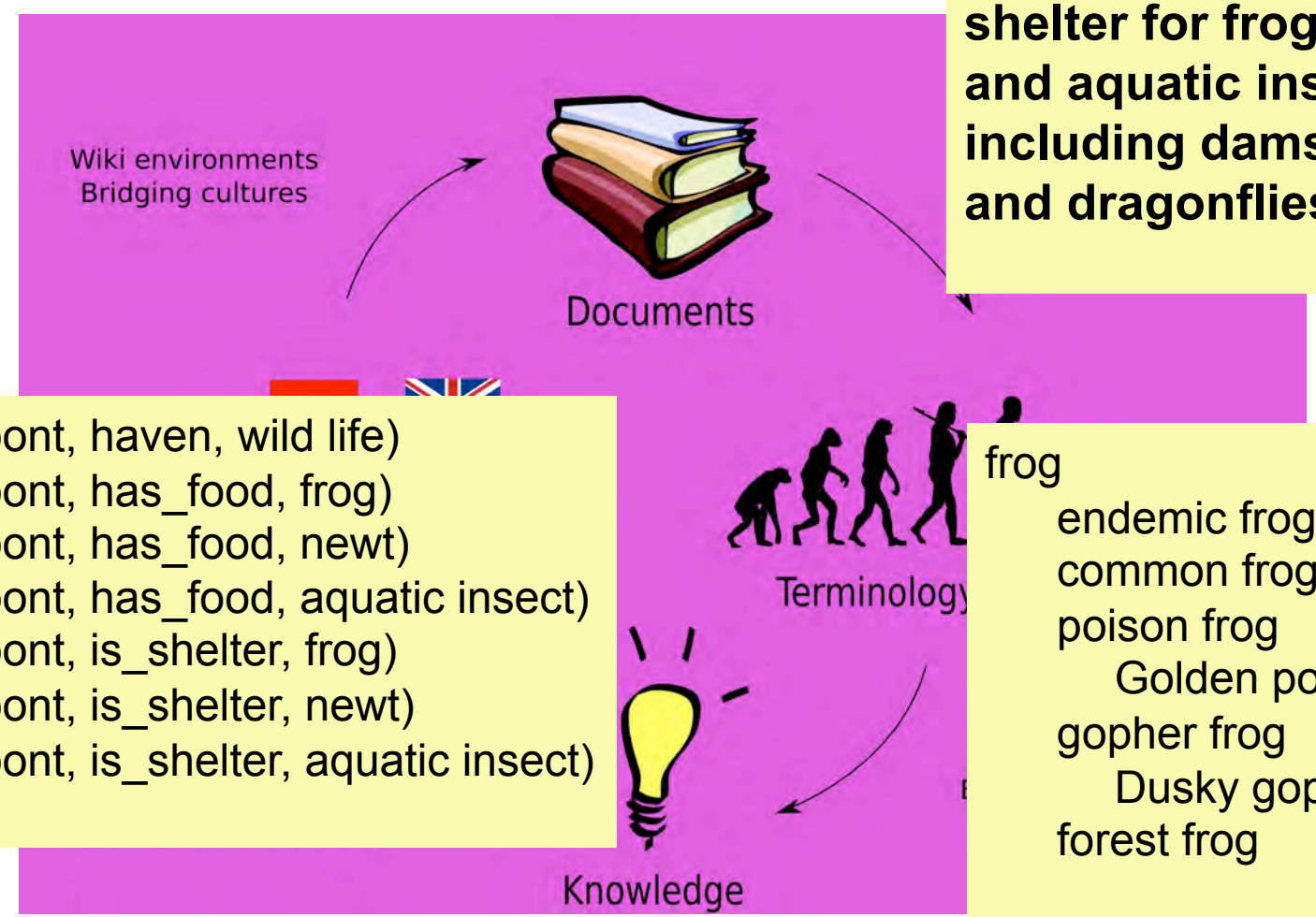
- **Languages**:
  - English, Dutch, Italian, Spanish, Basque, Chinese, Japanese
- **Domain**:
  - Environmental domain, BUT usable in any domain
- **Global:**
  - Both European and non-European languages
- **Available**:
  - Free: as open source system and data (GPL)
- **Future perspective**:
  - Content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

Current situation

- Vast amount of information in all kinds of formats and structures: websites, documents, databases, experts, community networks
- Scattered over the world: different regions, languages and cultures
- Highly dynamic and developing
- Involving an extreme range of interests and expertise: *government, general public, education, policies, regulations and rules, biology, health, chemistry, agriculture, economy, poverty, social impacts, transport, tourism, food industry, natural disasters, etc*.

KYOTO cycle

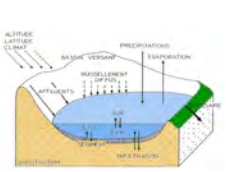
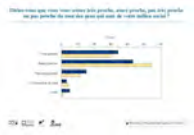
Garden ponds are havens for wildlife. They provide food and shelter for frogs, newts and aquatic insects, including damselflies and dragonflies,



(garden pont, haven, wild life)
(garden pont, has_food, frog)
(garden pont, has_food, newt)
(garden pont, has_food, aquatic insect)
(garden pont, is_shelter, frog)
(garden pont, is_shelter, newt)
(garden pont, is_shelter, aquatic insect)

frog
- endemic frogs
- common frog
- poison frog
  - Golden poison frog
- gopher frog
  - Dusky gopher frog
- forest frog

Distributed, diverse & dynamic data

1

Environmental organizations

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

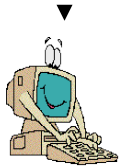
H20 CO2

H20 Pollution | CO2 Emission | Greenhouse Gas

Top

Middle

Domain



- **Tybot: term yielding robot**

3

CO2 emission

- **Kybot: knowledge yielding robot**

5

Index facts:

Process:	Increase
Involves:	CO2 emission
When:	2008
Where:	Europe

Text & Fact Index

6

Semantic Search

Kyoto main assets

- Wiki platform (WIKYOTO) for connecting, transferring and controlling knowledge and information across people and computers
- Term yielding robots (TYBOT): software that extracts terms and concepts from documents
- Knowledge yielding robots (KYBOT): fact extraction software that generates a comprehensive list of facts from collection of sources
- Fact alert: reports changes in facts on a collection of sources

What makes KYOTO unique?

- Wikyoto community tool that hides technology and complex knowledge and language representation
- Operated by community people and not by knowledge engineers and language technology people
- Exploits massive labor force of communities all over the world
- Connects domain knowledge to generic knowledge and domain language to general language

Natural language representation

- Linguistic processors in each language generate a common output format, so-called Kyoto Annotation Format (KAF)
- Wordnets in all languages are represented in the same format: Wordnet-LMF
- Term extraction and Text mining is applied in the same way across the different languages

Role of Asian partners

- Validate that the annotation format also works for Chinese and Japanese -> done!
- Validate that Chinese and Japanese wordnets can be represented in Wordnet-LMF and mapped to English and a shared ontology -> done!
- Show that semantic search can also be applied from European to Asian languages and vice versa -> in progress
- Contribute to the discussion on cultural and conceptual differences across languages and cultures

Experiences positive so far

- Japanese partner is funded by funds of NICT, Kyoto;
- Communication is good and they deliver in time;
- They are not the leading partner and demands are lower;
- High degree of technology sharing and exchange in the project;
- Their role is crucial for the ambition to standardize the interpretation of language