Multilinguality in eCommerce
Knowledge-Based Mediation

LangTech 2003, Paris, France
24 November 2003

Johannes Heinecke, Alain Léger France Telecom R&D
Aarno Lehtola, Cathérine Bounsaythip, VTT
Context

Multi-lingual Knowledge Based European Electronic Marketplace

QIST-FP5, funded by the European Commission
Industrial and academic partners under the coordination of France Telecom R&D
- VTT Information Technology (Finland)
- National Technical University of Athens (Greece)
- CNRS-LIRMM (France)
- SNCF (France)
- SchlumbergerSema (Spain)
- Ellos (Finland)
- Fidal (France)
- Universidad Politécnica de Madrid (Spain)

Q http://www.mkbeem.com/
Online Language Challenges for eCommerce

Native English speakers comprise less than 9% of the world population

“If I’m selling to you, I speak your language. If I’m buying, dann müssen Sie Deutsch sprechen”
(Willy Brandt)

<table>
<thead>
<tr>
<th>Year</th>
<th>English</th>
<th>Non-English</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>40.0</td>
<td>10.0</td>
<td>50.0</td>
</tr>
<tr>
<td>1997</td>
<td>45.0</td>
<td>16.0</td>
<td>61.0</td>
</tr>
<tr>
<td>1998</td>
<td>72.0</td>
<td>45.3</td>
<td>117.0</td>
</tr>
<tr>
<td>1999</td>
<td>148.0</td>
<td>209.1</td>
<td>357.1</td>
</tr>
<tr>
<td>2000</td>
<td>192.0</td>
<td>231.0</td>
<td>423.0</td>
</tr>
<tr>
<td>2001</td>
<td>231.0</td>
<td>270.0</td>
<td>501.0</td>
</tr>
<tr>
<td>2002</td>
<td>245.0</td>
<td>340.0</td>
<td>585.0</td>
</tr>
<tr>
<td>2003</td>
<td>270.0</td>
<td>408.0</td>
<td>678.0</td>
</tr>
<tr>
<td>2004</td>
<td>295.0</td>
<td>480.0</td>
<td>775.0</td>
</tr>
<tr>
<td>2005</td>
<td>320.0</td>
<td>540.0</td>
<td>860.0</td>
</tr>
</tbody>
</table>

Ref: Global Reach: http://www.glreach.com/

Distribution of this document is subject to France Telecom’s authorization
LangTech 2003 Heinecke/Léger/Lehtola/Bounsaythip 3 - 24/11/2003
MKBEEM - Multilingual eCommerce

The goal is to develop intelligent knowledge-based key components (HLP & KRR) for applications in multilingual eCommerce.

- Language adaptation via automatic HL translation and interpretation
- Natural dialogues combining HL and navigation
- Harmonised ontologies enabling localised views to products and trading contracts

Customer language information retrieval & trading

Multilingual cataloguing: write once, publish many

Transactions with contract adaptation

CP/SP User

Monolingual CP/SP

CP/SP eCom Service

Generic solutions proved by trials in Finnish, French and English in the domains of travel and mail-order sales

France Telecom R&D
Objectives

The global aim
- to extend current electronic commerce platforms
  - to reach a European and culturally open electronic commerce market.
- main technical aim of MKBEEM
  - to create an intelligent knowledge based multilingual mediation service
  - Natural language interfaces for both the system’s content providers/service providers and the end user.
  - Automatic multilingual cataloguing of products by service providers.
  - On-line e-commerce contractual negotiation mechanisms in the language of the user, which guarantee safety and freedom.

Domains
- mail order (clothing)
- train reservation, hotel/accommodation reservation, car rental
Language Independant Ontologies

Cataloguing of Product Descriptions

Multilingual Product Data


"Quilted jacket. Puffy model with reinforcements on the shoulder..."

data(Node) jcket(X,quilted_jacket), model(X,puffy), part(X,Y,sleeves), property(Y,Z,reinforcement)

Meaning extraction

Machine translation

Dialogue processing

HL processing

Material Ontology

Colour Ontology

Product Model

User Information request

A brown jacket made of natural material

14 products found:
1. Beige winterjacket of wool
2. Ochre quilted jacket of cotton
...

Any further requirements?

One with a hood

France Telecom R&D

Distribution of this document is subject to France Telecom’s authorization
LangTech 2003 Heinecke/Léger/Lehtola/Bounsaythip 6 -24/11/2003
Language Independant Ontologies (II)
Multilingual Cataloguing Tool of Mkbeem

The new product belongs to the supported product domains.

Available a picture and a textual product description in one of the supported languages.

France Telecom R&D
Multilingual Cataloguing

- Importing/editing/deleting product articles
- Translation of product articles to all target languages
- Text Checking to validate the input
- Extracting product properties from product articles
- Finding Categories to assign to the product in each market
- Search to locate pre-existing product descriptions for maintenance

France Telecom R&D
Observed benefits:

- Helps in producing consistent and uniform information
- Can make the working process faster and reduce the amount of manual, repeated routine procedures
- Cataloguing process as a whole was seen as an easy and efficient way of producing and classifying product information
- Very important: semi-automatic translation into target languages (post-editing possible)
- Property extraction and inference with colours and materials seen as important in bringing value-adding services to customers
- Product model and lexicon management tools considered suitable to their task

Reported difficulty:

- Occasionally long response times → frustration of the user
  - e.g. “hourglass” or provision of partial results could bring quick help
  - Will be eventually solved by continued product development, the current implementation based on Enterprise Java Beans and Applets
Market Perspectives for the Multilingual Cataloguing Tool

A portal operator could provide the tool for particular domains:

- This could be an easy solution for domains with large number of small or even micro businesses involved, like renting of vacation cottages.

The tool can be embedded into the catalogue production process of a seller company:

- It likely changes the organisation of work and requires business process re-engineering.
- The cataloguing tool can not be considered as package software. Implementing it to production use requires adaptation and tailoring.
Natural Language Request Analysis

S Four basic services:
  Q train reservation, accommodation reservation, car rental and mail ordering
  Q Combination of several services possible:
    – I want to visit Paris and reserve a hotel next weekend

S Several steps to create the correct ontological formula
  Q Language identification (English, French, Spanish or Finnish)
  Q Syntactical parsing and creating of a semantic representation
    – dependency trees are used to build DRT like structures (graphs)
  Q Checking against the linguistic ontology
    – inappropriate graphs are deleted
    – treatment of temporal expressions
  Q Creating of the final ontological formula using the concepts defined in the main ontology (internal format, OWL compatible)
Processing of temporal deictics

Transformation into corresponding absolute temporal expressions:

- **deictic elements**
  - *now, today, in two hours, in five days, next Monday, at ten to eleven pm*

- **incomplete or varying dates**
  - *the 12th of April, on Good Friday*

<table>
<thead>
<tr>
<th>Time Expression</th>
<th>Absolute Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>now</td>
<td>20.10.2003 13:56</td>
</tr>
<tr>
<td>today</td>
<td>20.10.2003</td>
</tr>
<tr>
<td>in two hours</td>
<td>20.10.2003 15:56</td>
</tr>
<tr>
<td>in five days</td>
<td>25.10.2003</td>
</tr>
<tr>
<td>next Monday</td>
<td>27.10.2003</td>
</tr>
<tr>
<td>at ten to eleven pm</td>
<td>20.10.2003 22:50</td>
</tr>
<tr>
<td>the 12th of April</td>
<td>12.04.2004</td>
</tr>
<tr>
<td>on Good Friday</td>
<td>9.04.2004</td>
</tr>
</tbody>
</table>
Example I

Phrase:
Q“I’ll arrive in Paris on Monday evening and I look for an accommodation with swimming pool.”

Semantic representation:
coord(coord1=x3005, coord2=x3006) &
arrival(destination=x3009, origin=x3010, situation=x3005, agent=x3013) &
speaker(theme=x3013) &
Paris(town=x3015, location=x3009) &
weekday~monday(date=x3005, wday=x3014) &
monthday~27(date=x3005, day=x3069) &
month~october(date=x3005, month=x3070) &
year~2003(date=x3005, year=x3071) &
hour~18(time=x3005, hour=x3072) &
minute~0(time=x3005, minute=x3073) &
staying(agent=x3021, situation=x3006, place=x3022, means=x3023, leisure=x3024) &
speaker(theme=x3021) &
accomodationorg(city=x3022, theme=x3023, leisure=x3024) &
swimmingPool(type=x3024).
Example II

Transformation into the ontological formula

\[(\text{trip})(V5609),\]
\[(\text{arrPlace})(V5609, \text{properName_Paris}),\]
\[(\text{date})(C63),\]
\[(\text{weekday})(C63, \text{monday}),\]
\[(\text{day})(C63, 27),\]
\[(\text{month})(C63, \text{october}),\]
\[(\text{year})(C63, 2003),\]
\[(\text{arrDate})(V5609, C63),\]
\[(\text{time})(C64),\]
\[(\text{hour})(C64, 18),\]
\[(\text{minute})(C64, 0),\]
\[(\text{arrTime})(V5609, C64),\]
\[(\text{accommodation})(V5610),\]
\[(\text{leisure})(V5610, \text{swimmingPool})\]
**Architecture of the MKBEEM ontology**

- **Service level:**
  - Generic services: provider-independent predefined offers

- **Global ontology:**
  - (Describes the common terms used in the whole MKBEEM platform (knowledge reusable on different domains)

- **Domain ontologies:**
  - Contain concepts corresponding to a specific domain (e.g., tourism, mail orders, etc.)

- **Source level:**
  - Specify the providers competencies
Formalisation of the best covering problem

**Notion of “best cover”**

- **$Q = \text{msc}(OF)$**
- **$E := \bigcap S_j$ (Conjunction of e-services)**
- **$\text{Rest}_E(OF) = Q - \text{lcs}(Q,E)$**
- **$\text{Miss}_E(OF) = E - \text{lcs}(E,Q)$**

**Definitions**

- **msc**: most specific concept
- **lcs**: least common subsumer

*France Telecom R&D*
# Example of e-service discovery

## RestE (OF)

- (trip)(V601)
- (arrPlace)(V601, properName_Paris)
- (date)(C32)
- (weekday)(C32, monday)
- (day)(C32, 27)
- (month)(C32, october)
- (year)(C32, 2003)
- (hotel)(V600)
- (leisure)(V600, swimmingPool)

## MissE (OF)

- (of)

## Extract of the e-service ontology:

### defconcept trip

\[
(\text{and} \ (\text{ATMOST} \ 1 \ \text{depPlace}) \ (\text{ATLEAST} \ 1 \ \text{depPlace}) \ (\text{ALL} \ \text{depPlace} \ \text{string}) \\
(\text{ATMOST} \ 1 \ \text{arrPlace}) \ (\text{ATLEAST} \ 1 \ \text{arrPlace}) \ (\text{ALL} \ \text{arrPlace} \ \text{string}) \\
(\text{ATLEAST} \ 1 \ \text{transportMean}) \ (\text{ALL} \ \text{transportMean} \ \text{transportMeanType})))
\]

### defconcept singleTicket

\[
(\text{and} \ \text{trip} \ \text{oneway})
\]

### defconcept returnTicket

\[
(\text{and} \ \text{trip} \ \text{twoways})
\]

### defconcept timetableSingle2

\[
(\text{and} \ \text{eserviceTrain4} \ (\text{ALL} \ \text{transportMean} \ \text{train}) \\
\text{singleTicket} \ (\text{ALL} \ \text{arrTime} \ \text{time}) \ (\text{ATLEAST} \ 1 \ \text{arrTime}) \ (\text{ATMOST} \ 1 \ \text{arrTime}) \\
(\text{ATMOST} \ 1 \ \text{arrDate}) \ (\text{ATLEAST} \ 1 \ \text{arrDate}) \ (\text{ALL} \ \text{arrDate} \ \text{date}))
\]

### defconcept accommodation

\[
(\text{and} \ (\text{ATMOST} \ 1 \ \text{placedIn}) \ (\text{ATLEAST} \ 1 \ \text{placedIn}) \ (\text{ALL} \ \text{placedIn} \ \text{string}) \\
(\text{ATMOST} \ 1 \ \text{startDate}) \ (\text{ATLEAST} \ 1 \ \text{startDate}) \ (\text{ALL} \ \text{startDate} \ \text{date}) \\
\text{user} \ (\text{ATMOST} \ 1 \ \text{room_type}) \ (\text{ATLEAST} \ 1 \ \text{room_type}) \ (\text{ALL} \ \text{room_type} \ \text{discrete}) \\
(\text{ATMOST} \ 1 \ \text{numberOfNight}) \ (\text{ATLEAST} \ 1 \ \text{numberOfNight}) \\
(\text{ALL} \ \text{numberOfNight} \ \text{number}) \\
(\text{ATMOST} \ 1 \ \text{numberOfClient}) \ (\text{ATLEAST} \ 1 \ \text{numberOfClient}) \\
(\text{ALL} \ \text{numberOfClient} \ \text{number}))
\]

### defconcept hotel

\[
(\text{AND} \ \text{eserviceHotel} \ \text{accommodation} \ (\text{ALL} \ \text{leisure} \ \text{discrete}) \ (\text{ALL} \ \text{numberOfBeds} \ \text{number}) \\
(\text{ALL} \ \text{hotelName} \ \text{discrete}))
\]
Example II

Reminder:
“I’ll arrive in Paris on Monday evening and I look for an accommodation with swimming pool.”

Extract from the service ontology (arrival date)

```
(defconcept hotel (AND eserviceHotel accommodation
    (ALL numberOfBeds number)
    (ALL hotelName string)))
(defconcept apartment (AND eserviceApartment accommodation
    (ATMOST 1 numberOfRooms)
    (ATLEAST 1 numberOfRooms)
    (ALL numberOfRooms number)
    (ATMOST 1 appartmentCategory)
    (ATLEAST 1 appartmentCategory)
    (ALL appartmentCategory string))
(defconcept timetableSingle1(AND eserviceTrain1 trip
    (ALL depTime time)(ATLEAST 1 depTime)
    (ATMOST 1 depTime)
    (ATMOST 1 depDate)(ATLEAST 1 depDate)
    (ALL depDate date)))
(defconcept timetableSingle2(AND eserviceTrain2 trip
    (ALL arrTime time)(ATLEAST 1 arrTime)
    (ATMOST 1 arrTime)
    (ATMOST 1 arrDate)(ATLEAST 1 arrDate)
    (ALL arrDate date)))
```

Identified services: timetable2, hotel
Missing information: numberOfBeds, hotelName

Identified services: timetable2, apartment
Missing information: numberOfRooms, appartmentCategory
Example III

Imagine a less specified user request:

“I’ll arrive in Paris and I look for an accommodation with swimming pool.”

<table>
<thead>
<tr>
<th>Identified services:</th>
<th>Identified services:</th>
</tr>
</thead>
<tbody>
<tr>
<td>timetableSingle1, hotel</td>
<td>timetableSingle1, appartment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Missing information:</th>
<th>Missing information:</th>
</tr>
</thead>
<tbody>
<tr>
<td>depTime.hour, depTime.minute</td>
<td></td>
</tr>
<tr>
<td>depDate.day, depDate.month</td>
<td></td>
</tr>
<tr>
<td>depDate.year, depDate.weekday</td>
<td></td>
</tr>
<tr>
<td>numberOfBeds, hotelName</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identified services:</th>
<th>Identified services:</th>
</tr>
</thead>
<tbody>
<tr>
<td>timetableSingle2, hotel</td>
<td>timetableSingle2, appartment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Missing information:</th>
<th>Missing information:</th>
</tr>
</thead>
<tbody>
<tr>
<td>arrTime.hour, arrTime.minute</td>
<td></td>
</tr>
<tr>
<td>arrDate.day, arrDate.month</td>
<td></td>
</tr>
<tr>
<td>arrDate.year, arrDate.weekday</td>
<td></td>
</tr>
<tr>
<td>numberOfBeds, hotelName</td>
<td>numberOfRooms, appartmentCategory</td>
</tr>
</tbody>
</table>
Conclusion of Results

- Successful implementation of multilingual mediation system, based on knowledge, coded in ontologies.

- Performing:
  - Language identification,
  - Semantic analysis of user request,
  - Transformation into an language independent ontological formula

- Identifying the service/product the user wants to buy by the help of service ontologies

- Existing parameters are extracted, missing ones requested in a subsequent step.

- Data base of the appropriate content provider is contacted

- The user is presented the results of his initial requests.
Thank you for your attention!

Contacts:

Qalain(dot)leger(at)rd(dot)francetelecom(dot)com
Qaarno(dot)lehtola(at)vtt(dot)fi
Qcatherine(dot)bounsaythip(at)vtt(dot)fi
QJohannes(dot)heinecke(at)rd(dot)francetelecom(dot)com
Future Development Recommendations

Further development of could focus on the following issues:

- Information request processing dialogues:
  - question answering capabilities (e.g. qualitative questions about the goods selection)
  - proper way of handling null queries (e.g. graceful relaxation of the search constraints based on the ontology models and the actual goods selection)

- New languages to the system: Russian, Norwegian, Estonian, German...

- User-friendlier ways for the acquisition and maintenance of language models and product models (knowledge acquisition bottleneck):
  - machine learning

- Adaptation to environments with limited resources, e.g. mobile terminals, with automatic text abstraction etc...
Work Needed for Adding New Product Domains and Languages

Marginal cost of adding a new domain or a new language is reasonable with respect to the added-value gained.

Based on experiences from modelling vacation cottage domain to the system (fi, fr, en) we have estimated that introducing a comparable new domain would require:
- semantic-lexicon: 2 person-months
- translation and meaning extraction rules: 1 person-month
- product models: 2-4 person-weeks

We also estimate that adding a language to a pre-existing domain would need:
- semantic-lexicon: 1-2 person-months
- translation and meaning extraction rules: 2-4 person-week
- product models: 1 person-week

Similar values for the tourism domain (train/hotel reservation and car rental)