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

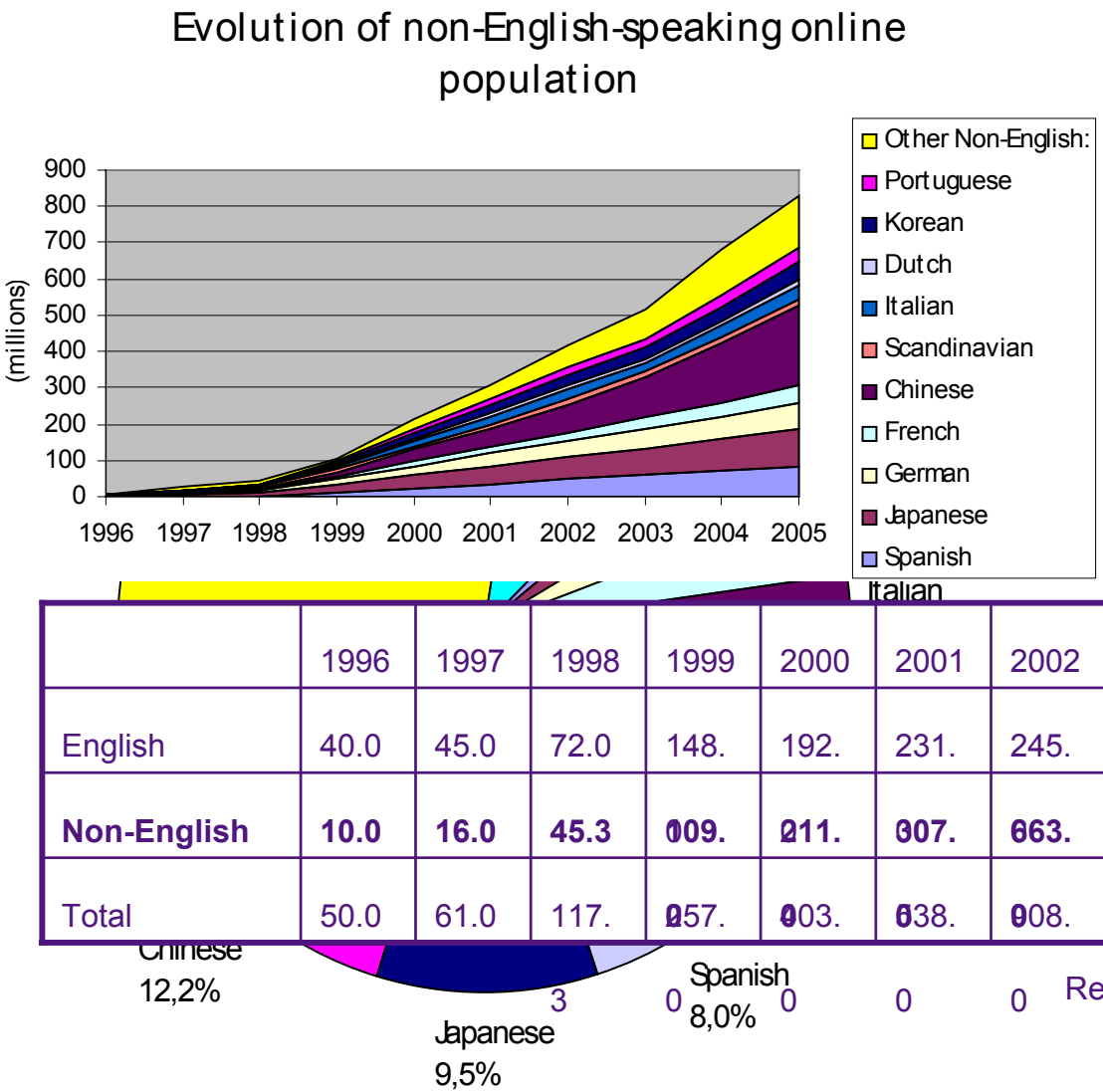
The present document contains information that remains the property of France Telecom. The recipient's acceptance of this document implies his or her acknowledgement of the confidential nature of its contents and his or her obligation not to reproduce, transmit to a third party, disclose or use for commercial purposes any of its contents whatsoever without France Telecom's prior written agreement.

Context

- Multi-lingual Knowledge Based European Electronic Marketplace
  - IST-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)
  - http://www.mkbeem.com/

Distribution of this document is subject to France Telecom's authorization

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)

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	Upper limit
English	40.0	45.0	72.0	148.	192.	231.	245.	270.	295.	320.0	555.0
Non-English	**10.0**	**16.0**	**45.3**	**109.**	**211.**	**307.**	**663.**	**640.**	**680.**	**820.0**	**5850.0**
Total	50.0	61.0	117.	257.	403.	538.	908.	910.	975.	1140.0	6400.0

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

MKBEEM - Multilingual eCommerce

Customer

Customer language information retrieval & trading

MKBEEM Mediation System

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

Multilingual cataloguing: write once, publish many

Transactions with contract adaptation

Monolingual CP/SP

CP/SP User

CP/SP eCom Service

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

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

( )

France Telecom R&D

Distribution of this document is subject to France Telecom's authorization

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

Distribution of this document is subject to France Telecom's authorization

Language Independant Ontologies

Cataloguing of Product Descriptions

Toppatakki. Muhkea malli, olkapäissä vahvikkeet. Painonapeilla kiinnitetty huppu, jossa joustava nyöri. Vetoketjun alla suojalista. Kaksi kannellista taskua...

HL processing

Meaning extraction

Machine translation

Dialogue processing

User Information request

A brown jacket made of natural material

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

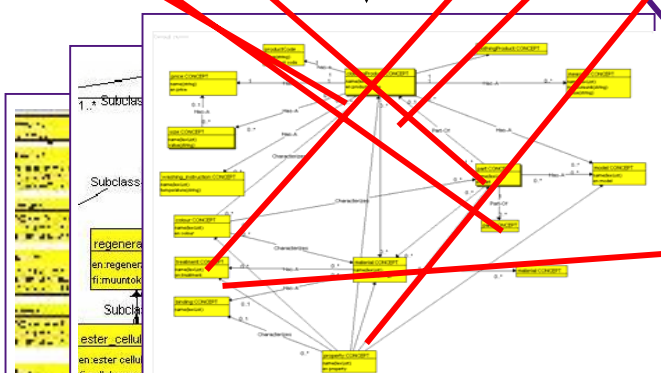
Any further requirements?

One with a hood

Multilingual Product Data

"Toppatakki. Muhkea malli..."
"Quilted jacket. Puffy model with reinforcements on the shoulder..."

jacket(X,quilted_jacket), model(X,puffy),
part(X,Y,sleeves),
property(Y,Z,reinforcement)



Product Model

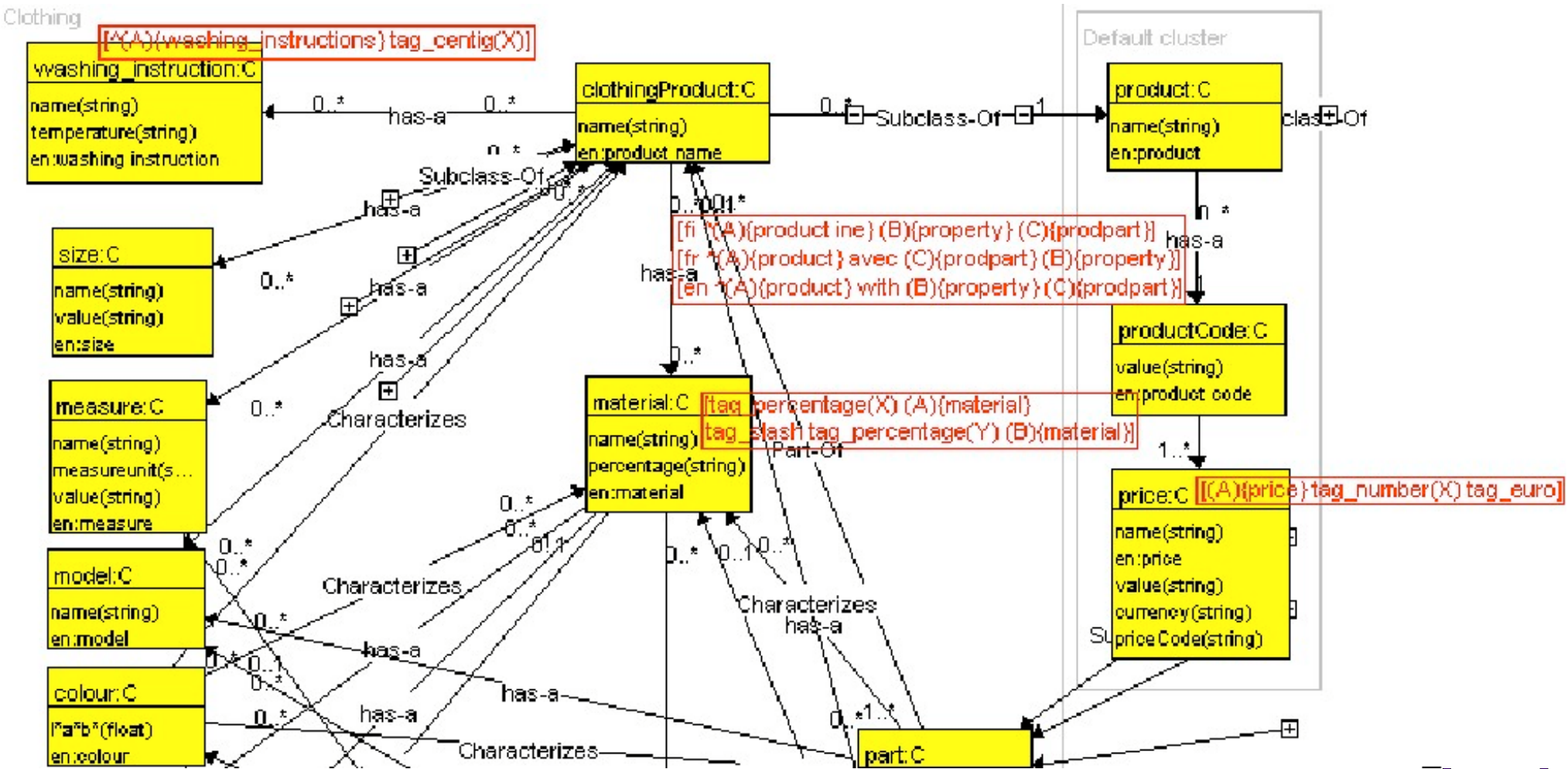
Material Ontology

Colour Ontology

France Telecom R&D

Distribution of this document is subject to France Telecom's authorization

Language Independant Ontologies (II)

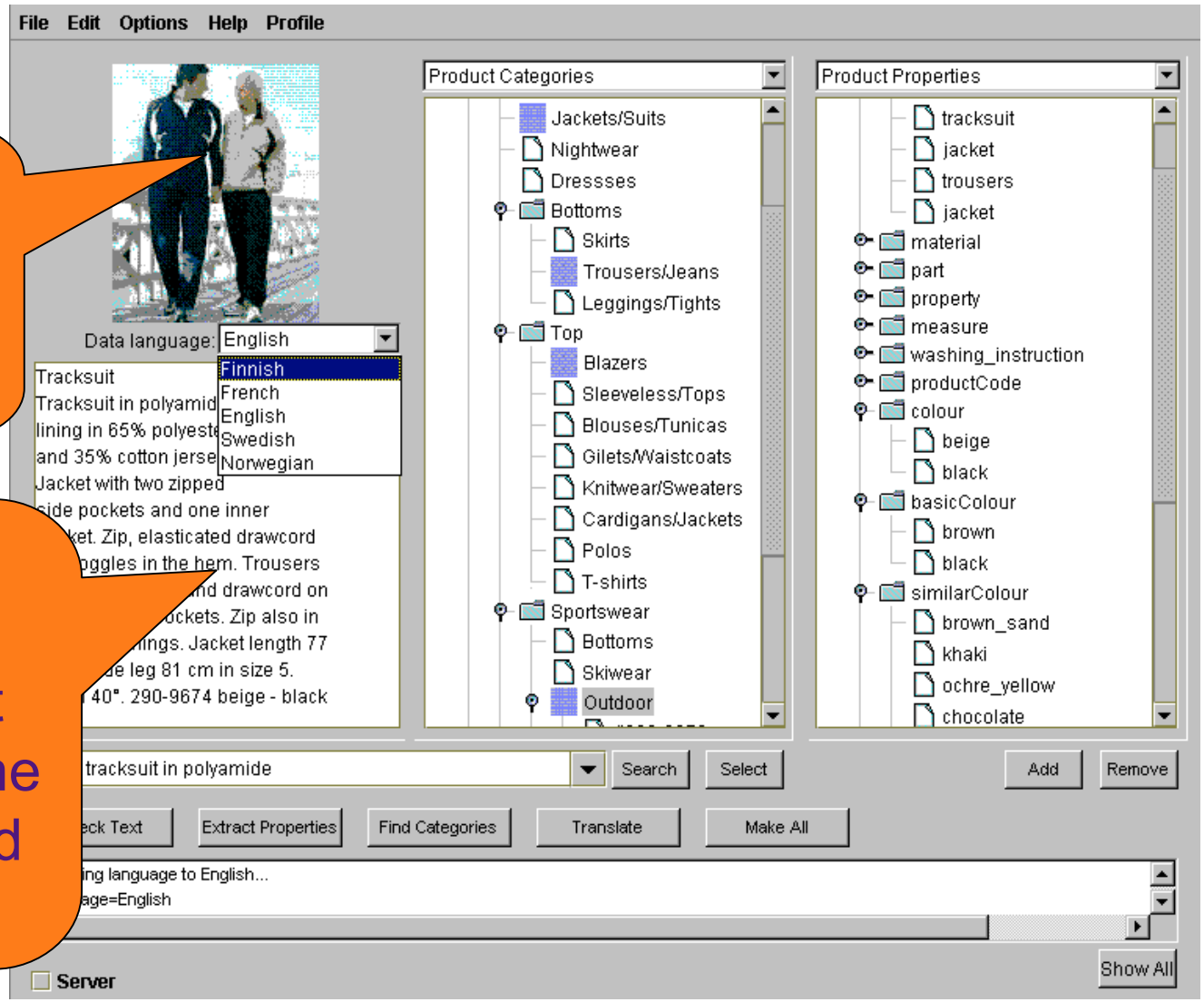


Distribution of this document is subject to France Telecom's authorization

Multilingual Cataloguing Tool of Mkbeem

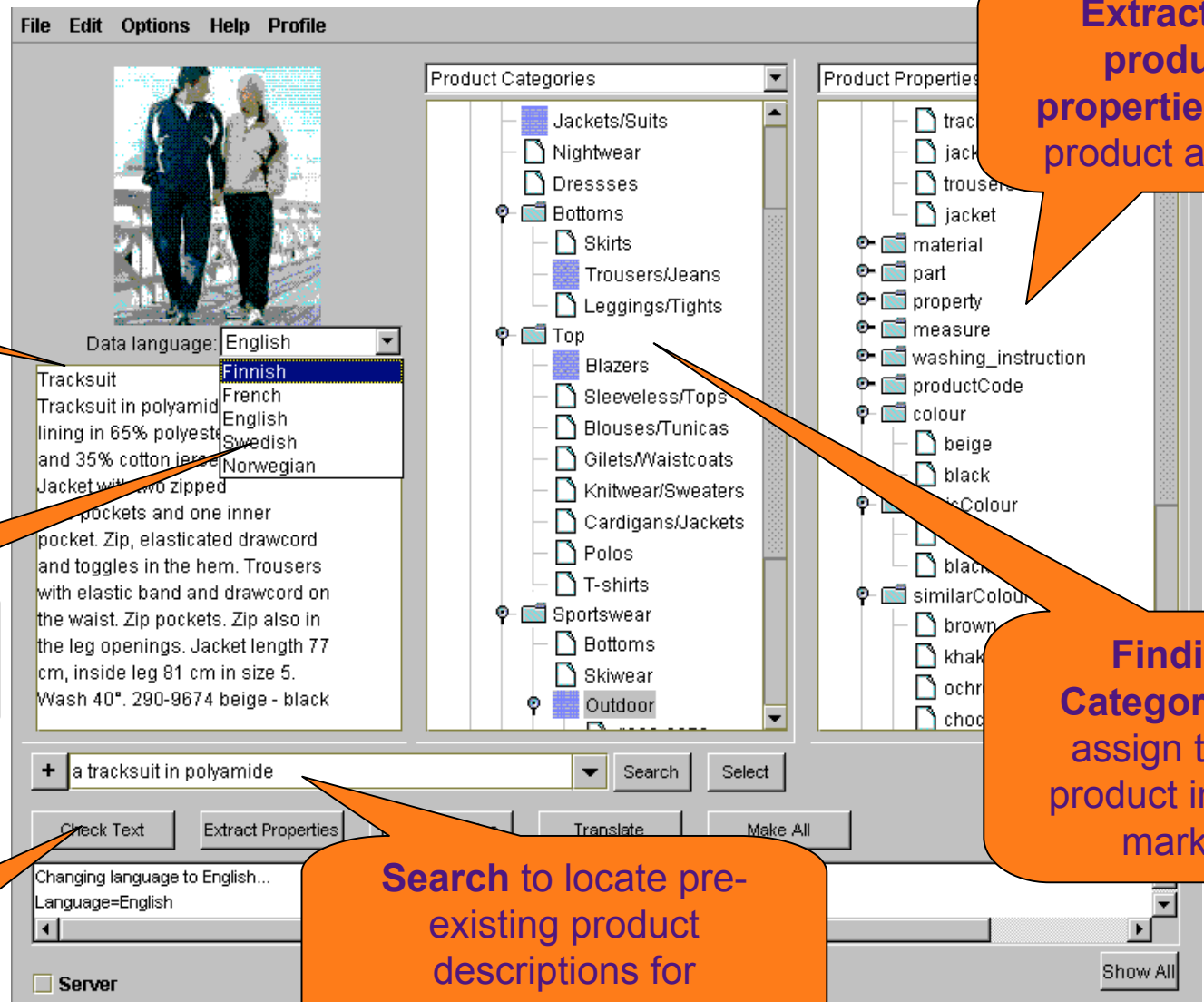
The new product belongs to the supported product domains

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



Distribution of this document is subject to France Telecom's authorization

Multilingual Cataloguing



Importing/ editing/deleting product articles

Translation of product articles to all target languages

Text Checking to validate the input

Search to locate pre-existing product descriptions for maintenance

Finding Categories to assign to the product in each market

Extracting product properties from product articles

bution of this document is subject to France Telecom's authorization

Ellos test group found the tool useful

- 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.

Distribution of this document is subject to France Telecom's authorization

Natural Language Request Analysis

- Four basic services:
  - train reservation, accommodation reservation, car rental and mail ordering
  - Combination of several services possible:
    - *I want to visit Paris and reserve a hotel next weekend*
- Several steps to create the correct ontological formula
  - Language identification (English, French, Spanish or Finnish)
  - Syntactical parsing and creating of a semantic representation
    - dependency trees are used to build DRT like structures (graphs)
  - Checking against the linguistic ontology
    - inappropriate graphs are deleted
    - treatment of temporal expressions
  - Creating of the final ontological formula using the concepts defined in the main ontology (internal format, OWL compatible)

Distribution of this document is subject to France Telecom's authorization

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*

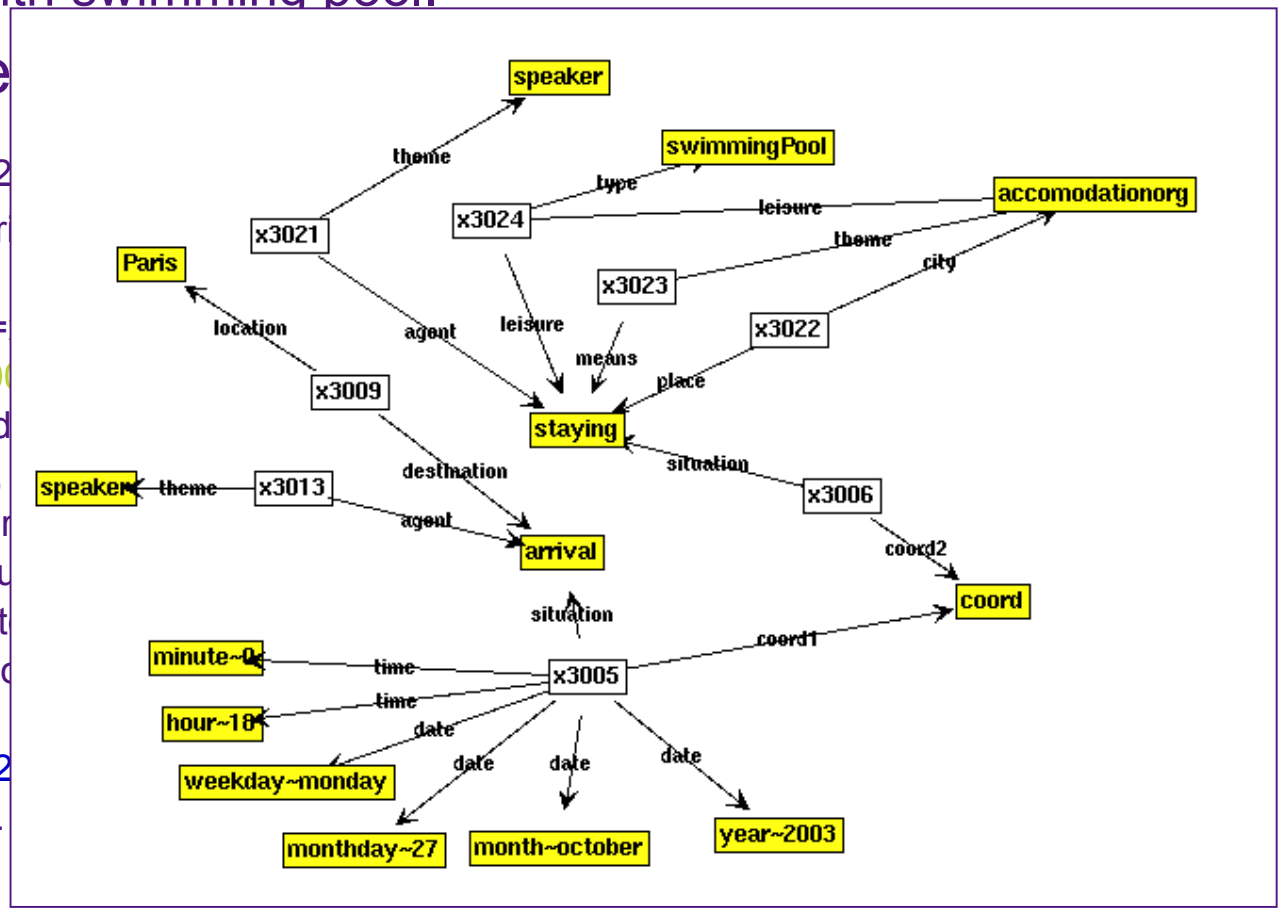
now	20.10.2003 13:56
today	20.10.2003
in two hours	20.10.2003 15:56
in five days	25.10.2003
next Monday	27.10.2003
at ten to eleven pm	20.10.2003 22:50
the 12th of April	12.04.2004
on Good Friday	9.04.2004

Distribution of this document is subject to France Telecom's authorization

Example I

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

coord(coord1=x3005, coord2
arrival(destination=x3009, or
speaker(theme=x3013) &
Paris(town=u3015, location=
weekday~monday(date=x30
monthday~27(date=x3005, d
month~october(date=x3005,
year~2003(date=x3005, year
hour~18(time=x3005, hour=u
minute~0(time=x3005, minut
staying(agent=x3021, situatio
speaker(theme=x3021) &
accomodationorg(city=x3022
swimmingPool(type=x3024).

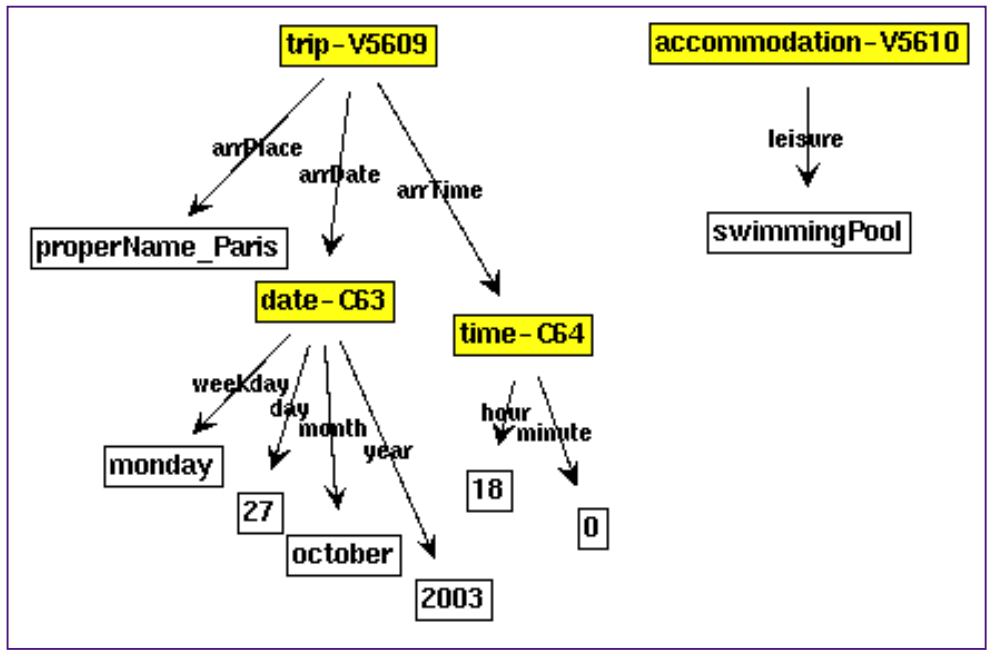


Distribution of this document is subject to France Telecom’s authorization

Example II

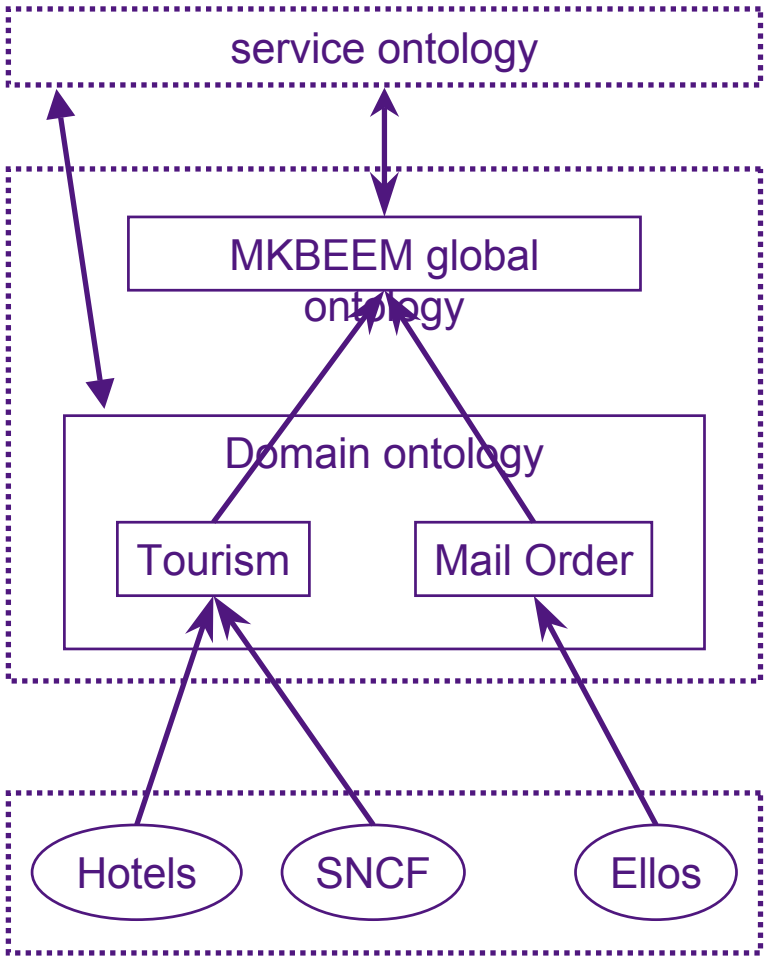
- Transformation into the ontological formula

(trip)(V5609),
    (arrPlace)(V5609, *properName_Paris*),
(date)(C63),
    (weekday)(C63, *monday*),
    (day)(C63, *27*),
    (month)(C63, *october*),
    (year)(C63, *2003*),
    (arrDate)(V5609, C63),
(time)(C64),
    (hour)(C64, *18*),
    (minute)(C64, *0*),
    (arrTime)(V5609, C64),
(accommodation)(V5610),
    (leisure)(V5610, *swimmingPool*)



Distribution of this document is subject to France Telecom's authorization

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.)

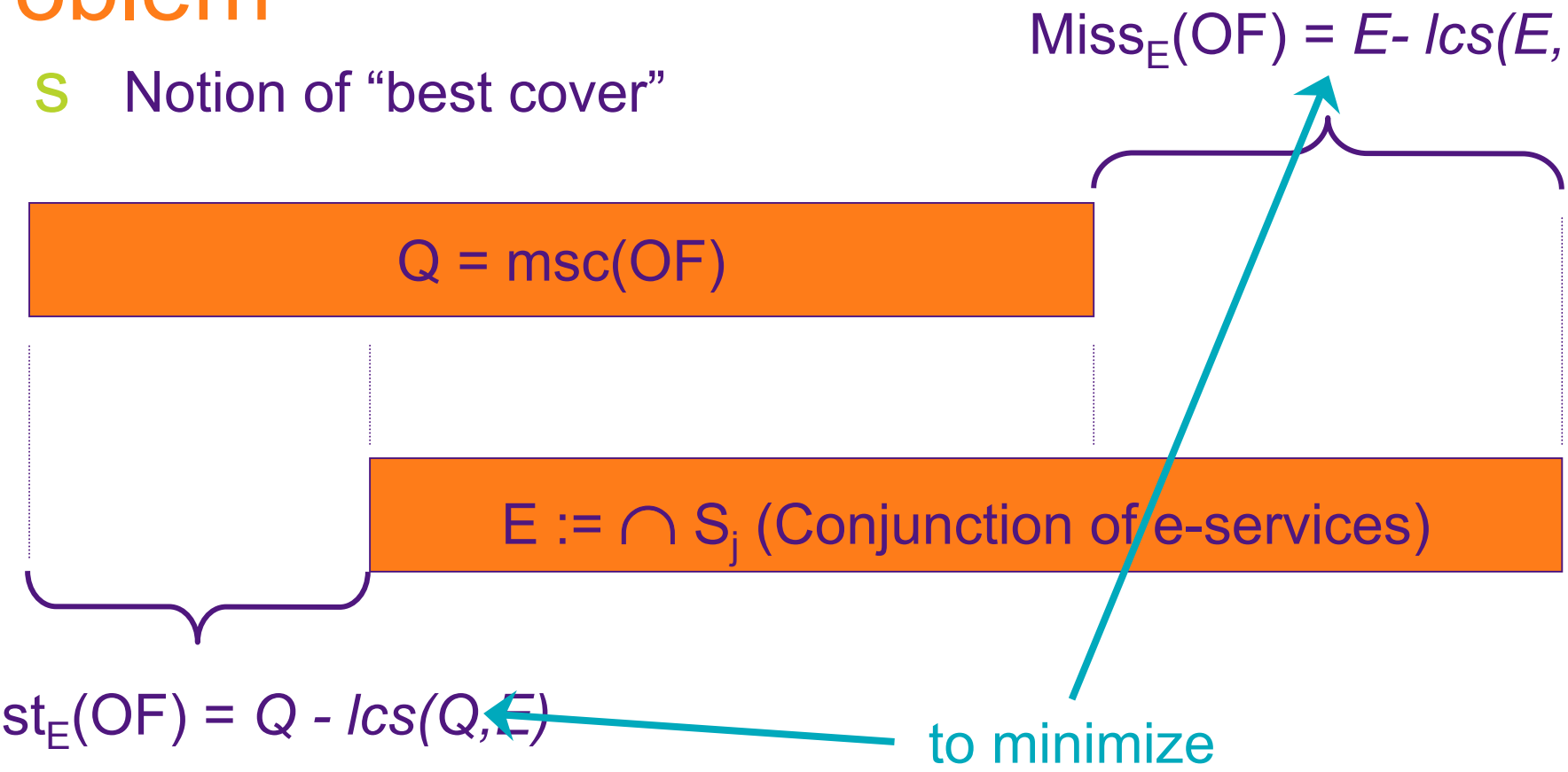
Source level:

Specify the providers competencies

Distribution of this document is subject to France Telecom's authorization

Formalisation of the best covering problem

- Notion of “best cover”

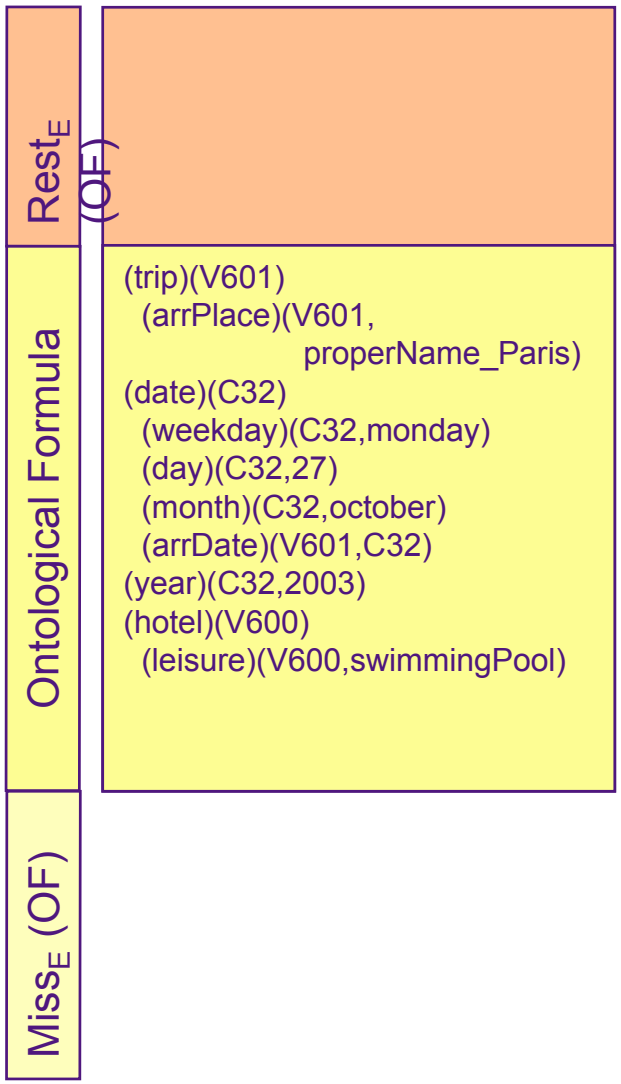


msc = most specific concept lcs = least common subsumer

( )

Distribution of this document is subject to France Telecom’s authorization

Example of e-service discovery



Extract of the e-service ontology:

```
defconcept trip
    (and (ATMOST 1 depPlace)  (ATLEAST 1 depPlace)  (ALL depPlace string)
         (ATMOST 1 arrPlace)  (ATLEAST 1 arrPlace)  (ALL arrPlace string)
         (ATLEAST 1 transportMean) (ALL transportMean transportMeanType)))

(defconcept singleTicket (and trip oneway))
(defconcept returnTicket (and trip twoways))

(defconcept timetableSingle2
   (and eserviceTrain4 (ALL transportMean train)
        singleTicket (ALL arrTime time)  (ATLEAST 1 arrTime) (ATMOST 1 arrTime)
                     (ATMOST 1 arrDate) (ATLEAST 1 arrDate) (ALL arrDate date)))
...
(defconcept accommodation
   (and (ATMOST 1 placedIn) (ATLEAST 1 placedIn) (ALL placedIn string)
        (ATMOST 1 startDate) (ATLEAST 1 startDate) (ALL startDate date)
        user  (ATMOST 1 room_type) (ATLEAST 1 room_type) (ALL room_type
discrete)
              (ATMOST 1 numberOfNight) (ATLEAST 1 numberOfNight)
              (ALL numberOfNight number)
              (ATMOST 1 numberOfClient) (ATLEAST 1 numberOfClient)
              (ALL numberOfClient number)))

(defconcept  hotel
            (AND eserviceHotel
                 accommodation  (ALL leisure discrete) (ALL numberOfBeds number)
                                (ALL hotelName discrete)))
```


Distribution of this document is subject to France Telecom's authorization

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 appartment (AND eserviceApartment accommodati
                            (ATMOST 1 numberOfRooms)
                            (ATLEAST 1 numberOfRooms)
                            (ALL numberOfRooms number)
                            (ATMOST 1 appartmentCategory)
                            (ATLEAST 1 appartmentCategory
                            (ALL appartmentCategory strin
(defconcept timetableSingle1 (AND eserviceTrain1 trip
                            (ALL depTime time)(ATLEAST 1
                            (ATMOST 1 depTime)
                            (ATMOST 1 depDate)(ATLEAST 1
                            (ALL depDate date)))
(defconcept timetableSingle2 (AND eserviceTrain2 trip
                            (ALL arrTime time)
                            (ATLEAST 1 arrTime)(ATMOST 1
                            (ATMOST 1 arrDate)(ATLEAST 1
                            (ALL arrDate date)))
```

Identified services:

timetable2, hotel

Missing information:

numberOfBeds, hotelName

Identified services:

timetable2, appartment

Missing information:

numberOfRooms,

appartmentCategory

Distribution of this document is subject to France Telecom's authorization

Example III

- Imagine a less specified user request:

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

Identified services: timetableSingle1, hotel Missing information: depTime.hour, depTime.minute depDate.day, depDate.month depDate.year, depDate.weekday numberOfBeds, hotelName	Identified services: timetableSingle1, appartment Missing information: depTime.hour, depTime.minute depDate.day, depDate.month depDate.year, depDate.weekday numberOfRooms, appartmentCategory
Identified services: timetableSingle2, hotel Missing information: arrTime.hour, arrTime.minute arrDate.day, arrDate.month arrDate.year, arrDate.weekday numberOfBeds, hotelName	Identified services: timetableSingle2, appartment Missing information: arrTime.hour, arrTime.minute arrDate.day, arrDate.month arrDate.year, arrDate.weekday numberOfRooms, appartmentCategory

Distribution of this document is subject to France Telecom’s authorization

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.

Distribution of this document is subject to France Telecom's authorization

Thank you for your attention!

- Contacts:
  - alain(dot)leger(at)rd(dot)francetelecom(dot)com
  - aarno(dot)lehtola(at)vtt(dot)fi
  - catherine(dot)bounsaythip(at)vtt(dot)fi
  - Johannes(dot)heinecke(at)rd(dot)francetelecom(dot)com

Distribution of this document is subject to France Telecom's authorization

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...

Distribution of this document is subject to France Telecom's authorization

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)

Distribution of this document is subject to France Telecom's authorization