Quarterly Management Reports for HLT projects under the IST programme

Project Ref. no.: IST-2000-25366  
Acronym: CROSSMARC  
Full Title: Cross-lingual Multi Agent Retail Comparison  
Reporting period: September - November 2001  
Project coordinator: Name: Constantine D. Spyropoulos  
Institution/Company: National Centre for Scientific Research “Demokritos”, Institute of Informatics & Telecommunications  
Address: 15310 Aghia Paraskevi Attikis, Athens, Greece  
Phone: +301-6503196  
Fax : +301-6532175  
E-mail: costass@iit.demokritos.gr  
Public project web site: http://www.iit.demokritos.gr/skel/crossmarc/

1. Progress

The 3rd project meeting (the minutes are attached), was held on November 15-16 in Rome with all partners attending, except ICN.

WP1. Collection of on-line product descriptions  
Deliverable1.1 “Report on the collection of on-line product descriptions for the 1st product type and on the techniques to access dynamically-generated pages” was sent to the EC. NCSR and VeltiNet developed the 1st version of the run-time collection tool (includes collection of product pages, XHTML conversion, demarcation). The 1st XML version of the 1st domain ontology was prepared by RTV.

WP2. Multilingual named-entity recognition and matching  
Deliverable D2.1 “Specification of Language Neutral Architecture for Named Entity Recognition” was sent to the EC. Corpus annotation task is about to complete using the NCSR annotation tool (EDIN-English, VeltiNet-Greek, NCSR-French, RTV-Italian). Draft versions of monolingual NERC have been prepared (EDIN-English, RTV-Italian, ICDC-French, NCSR-Greek) and presented during the Rome meeting.

WP3. Multilingual and multimedia fact extraction  
According to the TA, WP3 starts at month 10. However, the discussion on the FE DTD has started from month 9.

WP4: Prototype Development and Evaluation  
VeltiNet prepared and presented to the partners, during the Rome meeting, a proposal on the end-user interfaces, as well as a proposal on the architecture and the integration methodology.

WP5: Management and Dissemination  
The Deliverable D5.2 “Dissemination and Use Plan” was sent to the EC. CROSSMARC team members were present at relevant conferences, workshops and concertation events.
2. Deviations

There are no deviations in the workplan although there are still problems with ICN participation in the project. The tasks of ICN in WP1 were taken over by:
- NCSR and VeltiNet concerning the provision of French e-shops’ URLs to ICDC,
- NCSR concerning the task of annotating the corpus of French product pages provided by ICDC.

VeltiNet will also prepare a new workplan on WP4, proposing a re-allocation of the ICN tasks among the other partners.

3. Resources

<table>
<thead>
<tr>
<th>Participant's short name</th>
<th>WP1</th>
<th>WP2</th>
<th>WP3</th>
<th>WP4</th>
<th>WP5</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCSR</td>
<td>2.0</td>
<td>2.5</td>
<td></td>
<td>1.5</td>
<td>2.0</td>
<td>8.0</td>
</tr>
<tr>
<td>VeltiNet</td>
<td>1.5</td>
<td></td>
<td>1.5</td>
<td></td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>EDIN</td>
<td>0.3</td>
<td>2.0</td>
<td></td>
<td></td>
<td>0.3</td>
<td>2.6</td>
</tr>
<tr>
<td>RTV</td>
<td>1.0</td>
<td>5.0</td>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>8.0</td>
</tr>
<tr>
<td>ICDC</td>
<td>0.5</td>
<td>3.0</td>
<td>1.0</td>
<td></td>
<td>0.5</td>
<td>5.0</td>
</tr>
<tr>
<td>ICN1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>5.3</td>
<td>12.5</td>
<td>1.0</td>
<td>4.0</td>
<td>3.8</td>
<td>26.6</td>
</tr>
</tbody>
</table>

4. Cumulative

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Current status (completed/underway)</th>
<th>On schedule (yes/no)</th>
<th>Original completion date (as in Annex I &quot;Description of Work&quot; + Calendar Date, e.g. &quot;M06 = June 2000&quot;)</th>
<th>Actual/planned completion date (Calendar Date, e.g. &quot;15 June 2000&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.1</td>
<td>completed</td>
<td>Yes</td>
<td>M06=Aug 2001</td>
<td>Sept 2001</td>
</tr>
<tr>
<td>D2.1</td>
<td>completed</td>
<td>Yes</td>
<td>M06=Aug 2001</td>
<td>Sept 2001</td>
</tr>
<tr>
<td>D5.1</td>
<td>completed</td>
<td>Yes</td>
<td>M03=May 2001</td>
<td>June 2001</td>
</tr>
<tr>
<td>D5.2</td>
<td>completed</td>
<td>Yes</td>
<td>M06=Aug 2001</td>
<td>Sept 2001</td>
</tr>
<tr>
<td>1st Semestrial Report</td>
<td>completed</td>
<td>Yes</td>
<td>M06=Aug 2001</td>
<td>Sept 2001</td>
</tr>
<tr>
<td>1st Annual Report</td>
<td>completed</td>
<td>Yes</td>
<td>M09=Nov 2001</td>
<td>Jan 2001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant's short name</th>
<th>Old Total, carried over from last Management/Progress Report</th>
<th>This period's Total, carried over from 3. Resources</th>
<th>New TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCSR</td>
<td>15.3</td>
<td>8.0</td>
<td>23.3</td>
</tr>
<tr>
<td>VeltiNet</td>
<td>5.0</td>
<td>3.0</td>
<td>8.0</td>
</tr>
<tr>
<td>EDIN</td>
<td>9.5</td>
<td>2.6</td>
<td>12.1</td>
</tr>
<tr>
<td>RTV</td>
<td>13.0</td>
<td>8.0</td>
<td>21.0</td>
</tr>
<tr>
<td>ICDC</td>
<td>5.5</td>
<td>5.0</td>
<td>10.5</td>
</tr>
<tr>
<td>ICN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td><strong>48.3</strong></td>
<td><strong>26.6</strong></td>
<td><strong>74.9</strong></td>
</tr>
</tbody>
</table>

---

1  No figures have been given from ICN
Appendix:

*The minutes of the 3rd project meeting in Rome are attached.*
3rd CROSSMARC meeting,
Rome, Roma Tor Vergata
15-16 November 2001

Present:
--------
NCSR: Costas Spyropoulos, Vangelis Karkaletsis
VeltiNet: Mary Lagou
EDIN: Claire Grover
RTV: Maria Teresa Pazienza, Michele Vindigni, Fabio Zanzotto, Roberto Basili
ICDC: Francis Wolinski, Frantz Vichot

Work Package 1: Collection of on-line product descriptions

Current status
-------------------------------
- The 1st XML version of the 1st domain ontology was prepared by RTV
- VeltiNet and NCSR prepared the report on the domain typology and presentation layouts (it was included in D1.1)
- ICDC finalized the Corpus collection for the 1st domain using its tool. The corpus was sent to the partners for annotation.
- NCSR and VeltiNet Specs prepared the specs of the run-time collection tool (included in D1.1)
- Deliverable 1.1 was finalized and sent to the EC (NCSR, VeltiNet, ICDC, RTV)
- NCSR and VeltiNet developed the 1st version of the run-time collection tool (includes collection of product pages, XHTML conversion, demarcation)

Decisions and Actions
-------------------------------
1. RTV will finalise, by mid January, the XML version of the 1st domain ontology as well as the ontology maintenance process.
2. VeltiNet will coordinate the discussion on the specification of the 2nd domain. The final decision will be taken by the end of February (next meeting).
3. RTV and VeltiNet will prepare by the end of March the ontology and the template for the 2nd domain.
4. Partners (VeltiNet-Greek, NCSR-French, EDIN-English, RTV-Italian) will send URLs for the 2nd domain to ICDC, by mid April. ICDC will return the relevant pages by mid May.
5. NCSR, ICDC and VeltiNet will develop an improved version of the collection tool by end of February (next meeting).
6. NCSR will prepare a 1st draft report on learning techniques by end December setting the requirements for all languages.
7. **NCSR** coordinates preparation of Deliverable 1.2 “Report on the learning techniques”. This must be prepared by end of February.

8. **NCSR, ICDC and VeltiNet** will specify the presentation scenario for the review meeting by mid February.

**Work Package 2: Multilingual named-entity recognition and matching**

**Current status**

------------------------
- EDIN finalised the NERC DTD for the 1st domain
- Deliverable 2.1 was finalized and sent to the EC (EDIN, RTV, ICDC, NCSR).
- NCSR delivered the annotation tool and guidelines
- Corpus annotation task is about to complete using the annotation tool (EDIN-English, VeltiNet-Greek, NCSR-French, RTV-Italian)
- Draft versions of monolingual NERC have been prepared (EDIN-English, RTV-Italian, ICDC-French, NCSR-Greek) and presented during the meeting.

**Decisions and Actions**

-------------------------------
1. Partners (EDIN, RTV, ICDC, NCSR) should provide first evaluation results for monolingual NERC (recall, precision, F-measure) by mid January. The final evaluation results must be provided by mid February.
2. Partners (EDIN, RTV, ICDC, NCSR) must finalise the strategy for name matching and normalisation by end March (what can be performed before FE, evaluation strategy)
3. **NCSR** will prepare a new version of the annotation tool by end January, based on the partners’ remarks.
4. **EDIN** will coordinate the discussion on the NERC DTD for the 2nd domain. This must be finalised by end of April.
5. Annotation for the 2nd domain will start at the end of April (when relevant pages will be delivered by ICDC) and will be finalised by mid June. Partners involved are: **EDIN**-English, **VeltiNet**-Greek, **NCSR**-French, **RTV**-Italian.
6. **EDIN** will coordinate the preparation of Deliverable D2.2 (NERC v.1, system documentation). EDIN will send partners the deliverable template by mid. December. The 1st draft will be ready by mid January. The final version will be ready by end February.
7. Specify the presentation scenario for the review meeting by mid February (EDIN coordinates, ICDC checks Solaris / Linux).
8. 1st integrated version by next meeting (end February).

**Work Package 3: Multilingual and multimedia fact extraction**

**Current status**

-------------------------------
- NCSR distributed to partners a draft version of FE DTD for the 1st domain.

**Decisions and Actions**

-------------------------------
1. Partners decided on the features that will be included in the integrated FE prototype for the 1st domain (made_by, processor_name, processor_speed, soft_os, price).
2. **NCSR** will coordinate the discussion on the FE DTD. The new version, taking into account partners comments (esp. link to the ontology), will be sent by mid January along with an example of the XML file output by the FE module. The final FE DTD will be ready by end February. (next meeting)
3. **NCSR and Velti** will prepare a proposal and schedule on the corpus FE annotation task for the 1st domain by mid February.
4. Specify the presentation scenario for the review meeting by mid February
5. 1st integrated version by end March.

**Work Package 4: Prototype development and evaluation**

**Current status**
- VeltiNet presented a proposal on the end-user interfaces
- VeltiNet presented a proposal on the architecture and the integration methodology

**Decisions and Actions**
- **VeltiNet** coordinates the discussion on the architecture and integration methodologies, workplan, new version of the interfaces (country, vendor, user profile, language UI). A draft report on the above will be prepared by **VeltiNet, NCSR and EDIN** by mid December.
- **VeltiNet** coordinates the preparation of D4.1 (architecture of the integrated system and the separate monolingual systems, interfaces, black-box evaluation methodology, …). VeltiNet will send the deliverable template by mid December. A draft version will be ready by mid January. The final one will be ready by end February.
- 1st demo (INTERIM) prototype (incl. Web interface, named-entity recognition, 4 languages, 1 product type) by the end of February.

**Work Package 5: Management and Dissemination**

**Current status**
- Deliverables 5.1, 5.2 delivered to the EC (NCSR coordinated)
- 2nd Quarterly, 1st semestrial report and cost statements delivered to the EC (NCSR coordinated)
- NCSR handled ICN problem with the cooperation of the EC project officer (K.Rossi).
- Various dissemination activities (see 1st semestrial report)

**Decisions and Actions**
- Partners discussed the problem with ICN. Concerning ICN’ task of French URLs collection and annotation (both domains), NCSR undertakes it. Concerning ICN tasks in WP4, the responsible partner for WP4 (VeltiNet) will propose to the partners a re-allocation of work.
- Concerning the Consortium agreement, the partners decided to wait until the next meeting, since this is related to the ICN problem.
- NCSR will send a draft for the 1st Annual report by end November.
- NCSR will send a draft for the 3rd Quarterly report by end November.
- Partners discussed the letter of the Assist.Prof. Marios Dikaiakos, University of Cyprus, Dept. of Computer Science (www.ucy.ac.cy) concerning their request for participation in CROSSMARC (through the IST programme on extension of existing projects with partners from newly associated States). The main issue raised by the partners concerned the managerial complexities that may be caused from such extension. NCSR will contact EC in order to have a clear view on these issues and will inform the partners.
6. The partners discussed the next dissemination activities (folder, leaflets per language, participation in conferences, workshops). Decided to submit a paper to LREC’02. EDIN will issue a first draft.

7. Next meeting: **Athens, 28 Feb. – 01 March.**