

# **EVALITA 2009**

## **Spoken Dialogue System Task: Detailed Guidelines**

*Paolo Baggia, Francesco Cutugno, Morena Danieli, Roberto Pieraccini, Giuseppe Riccardi*

### **Introduction**

This document contains the guidelines for the participants of the Spoken Dialogue System Task (SDS, henceforth), part of the EVALITA 2009 evaluation campaign. Teams participating in this task will develop a voice application operating over telephonic channel. The application will model an interaction between an human caller and an automatic system for information retrieval operating in a specific domain.

### **Domain**

The SDS to be developed will operate in the sales force domain. The system should be able to serve salesmen calling their company and (1) asking for data about customers (customer identification code and/or customer personal data) or reviewing open orders/invoices, (2) requiring the opening of a new order of one or more positions, each of them including a product and a quantity; (3) searching the company catalog to find products and pricing and optionally discounts. Tasks (1 and 2) are considered as an ‘input’ procedure while task (3) is considered as ‘output’. Each call made to the SDS will be targeted to the completion of one or more of the three specific task listed above. A task will be considered successful if the input procedure will end with the correct recording of the provided information into the database and if the output process will deliver the correct prices of the correct products. Further details will be furnished on April the 1<sup>st</sup> (see following sections).

### **Database and Scenarios**

Starting from April the 1<sup>st</sup> 2009, the task coordinators will provide participants with an XML dump of a relational database conceptually structured as related entities (salesmen, customers, orders, products). Each table in the database will be populated with a sufficient number of items, the only accepted language for all contents will be Italian.

Scenarios:

Human-computer interaction should be designed to respond to the ‘mixed initiative’ requirements (as defined in Jurafsky & Martin 2009 pp. 829-831)

Each call will start with the identification of the caller as a salesman, this will be done either by providing a numeric ID as digit sequence, or as unique name.

After that the transaction should be oriented to the choice of the task to be conducted as first in the dialogue and can be modeled as menu driven.

After the initial choice, the following scenarios should be possible (the list is not exhaustive):

- a) **Q:** the salesman asks which orders are active for a specific customer.  
**A:** the system lists open orders, with positions, product and quantities.
- b) **Q:** the salesman records a new order related to selected customer, with one or more positions.  
**A:** the systems accepts the new orders.
- c) **Q:** the salesman consults the catalogue asking for products prices.  
**A:** the system reports prices referring discounts percentage according to order quantity.

## **Submission of System Results**

On September the 1<sup>st</sup> participants should provide a number of a fixed telephony line or an IP address supporting a SIP softphone connection to the coordination team.

Evaluation of the participating systems will be conducted in the period 3-20 September 2009 - by a panel of five experts that will directly call the system. Each expert will call each system at least 10 times.

Participant should record all calls in full duplex mode (minimum 8kHz sampling rate, mono, .wav format) and upload relative audio files into a repository whose URL will be successively published.

The SDS evaluation will take into account the successful completion of the scenarios assigned as well as the overall system performance and user experience.

## **Evaluation Metrics**

To evaluate systems both on-line and off-line methods will be used. Subjective and objective metrics will be adopted. You may find further details in:

<http://suendermann.com/su/pdf/slt2008a.pdf>

<http://www.disi.unitn.it/~riccardi/papers/sigdial08.pdf>

## **Contact Persons**

Morena Danieli: {morena.danieli@loquendo.com}

Francesco Cutugno: {cutugno@na.infn.it}

Roberto Pieraccini: {roberto@speechcycle.com}

Giuseppe Riccardi: {riccardi@disi.unitn.it}