

Evaluating Language Technology

EAGLES II Workshop on Evaluation Standards and Resource Sharing

by
Deborah Fry

It is only recently that human factors have become widely accepted as a requirement in the design and quality management of software systems as a whole. In an emerging area such as language engineering, with its many non-commercial and pre-industrial systems, the problem can be particularly acute. The work performed by the EAGLES II Evaluation Group on the ongoing development, discussion and dissemination of methodologies and metrics for language technology evaluation is thus a significant step on the road to industry growth. At a workshop held at the end of November 1997 in Brussels, the group presented its interim results and other key developments in the field to an international audience of participants from both the academic and commercial communities.

Originally launched in 1993 as part of the European Commission's Linguistic Research and Engineering (LRE) programme, EAGLES - the Expert Advisory Group on Language Engineering Standards - is designed to establish a consensus on a basis of facts and acceptable practices in a number of areas of language engineering. In this way, the project hopes to shorten the time which will be needed by the International Organization for Standardization (ISO) and other bodies to produce formal standards later on. In addition, it hopes to ensure greater integration of language technology applications and better re-use of existing technology, thus cutting development costs and supporting the growth of the market.

Uniting senior figures from major speech and language development projects across Europe, the project is now in its second stage, which will run until the second half of 1998. EAGLES II's job is to produce and disseminate manuals and guidelines on best practice and de facto standards, aimed at aiding both consumers and producers of language technology.

The project is divided into a number of different working groups, of which the Evaluation Group is one. As one of the ways of consolidating and disseminating its first-round work and of establishing an ongoing discussion forum for sharing results and experiences, the Group held an Evaluation Workshop at the end of November 1997 at the Free University of

Brussels. Maghi King of ISSCO (University of Geneva) presented the results produced by the first round of the Evaluation Group's work, which provides a coherent and pragmatic and widely accepted framework for language systems evaluation.

The new draft of the ISO 9126 standard, which defines quality characteristics for software product evaluation and provides a process model for their evaluation, was presented by its technical editor, Nigel Bevan of the National Physical Laboratory in the UK. Many of the ideas in this revised document are similar to those developed by the EAGLES group, raising the possibility of long-term harmonisation. (In fact, the EAGLES Group had taken the standard as one starting point in its work, but had felt that it did not cover all aspects of language engineering systems adequately at the time).

Another issue addressed at the Workshop was the ongoing work on defining a reference model for speech systems evaluation presented by Denis Johnston of British Telecom, which offered valuable insights into the long history of work in this area by the speech community and the definition and validation of evaluation metrics. Developer-oriented evaluation of system usability as represented by the Inuse and Megataq projects was outlined by Nigel Bevan and by Bert Arnold of Delft University respectively. Also on the agenda was a presentation by Karen Sparck-Jones (Cambridge University) on the US TREC program for evaluating retrieval systems, plus the idea of a "braided evaluation scheme". In this, the results of each stage in the evaluation process would provide the input for further work, with the aim of making optimum use of limited resources. The Workshop finished off with a discussion of evaluation activities and of the future focus of EAGLES work. This will include another workshop in 1998, building on the results of the first one and of the LREC Conference to be held in Granada in May 1998.

As an overview and discussion of the status of ongoing work in the field of language technology evaluation, the Brussels Workshop fulfilled an important role. The standard of discussion was high and the atmosphere lively and informal - something

that was reinforced by an excellent dinner on the first night. While much remains to be done on evaluation, the EAGLES Group can rightly be proud of its achievements. The participation of players from academic and business circles in both Europe and beyond is proof of the growing interest in the results being produced, and in the field in general. It is to be hoped that, as the language engineering industry continues to grow and gather speed, the pioneering work performed by the initiative will reach the much broader audience it deserves. ■

Deborah Fry is Co-director of Fry & Bonthrone Partnerschaft, a Language Services and Consultancy company based in Mainz-Kastel, Germany, and was one of the attendees at the Evaluation Workshop. She can be contacted via e-mail at deborah@mainz.netsurf.de or 100637.711@compuserve.com

Further information on the EAGLES initiative and its work is available from:

EAGLES II home page (includes information sharing pages and full Workshop Report)
<http://www.cst.ku.dk/projects/eagles2.html>

EAGLES I Final Report
<http://www.issco.unige.ch/egw95/>

General information on EAGLES
<http://www.ilc.pi.cnr.it/EAGLES96/>

EAGLES Evaluation Group Workshop

Maghi King,
ISSCO, University of Geneva
54 route des Acacias, CH-1227
Geneva, Switzerland
E-mail: maghi.king@issco.unige.ch

EAGLES
Co-Chief Editor
John McNaught,
Dept of Language Engineering,
UMIST, Sackville Street, PO Box 88
Manchester M60 1QD,
United Kingdom
E-mail: jock@ccl.umist.ac.uk