

Executive Summary

The aim of the MultiMatch evaluation activity will be to ensure that the individual components and the complete system prototypes are tested and that their development is assessed during the project life-cycle with respect to performance and usability (i.e. laboratory-based and user-centred evaluation). Component and system development will be monitored to ensure that the functional specifications defined in WP 1 are respected and performance evaluation will comprehend internal project assessment plus evaluation against external benchmarks. The usability testing will involve the MultiMatch user groups and a set of field trials will be organised.

This first deliverable of Workpackage 7 describes the approaches that will be used in order to evaluate the MultiMatch system. It focuses in particular on the definition of the methodology to be used for Prototype 1 both with respect to the laboratory-based and the user-centred evaluation; an update will be released at 24 months providing final details on the evaluation processes decided for Prototype 2. Deliverables 7.2 and 7.3, to be consigned at 18 and 29 months respectively, will describe the results of the evaluation of Prototypes 1 and 2 at component and overall system level. Deliverable 7.4, to be consigned at 30 months, will describe the results of the field trials with the operational system.

We adopt both laboratory-based and user-centred evaluation strategies in order to be able to assess as many aspects as possible of the system.

The former will allow us to evaluate component and overall system performance mainly from the perspective of retrieval accuracy. For the first prototype we will build a set of test collections to enable us to perform some project internal evaluation of the different system components, comparing results obtained using with various settings and configurations in order to be able to tune and improve performance. We will also duplicate some of the past experiments of CLEF - the Cross-Language Evaluation Forum¹ - so that we can compare performance against known benchmarks (as far as the availability of relevant data allows us). Our aim is to establish base-lines for the different components of the first prototype so that we can measure improvements achieved with the second prototype.

However, measures of retrieval accuracy do not always correlate well with user satisfaction and overall system performance. The user-centred evaluations will thus aim at measuring the level of user satisfaction with the system and will focus on assessing the various user interfaces and the overall system performance in terms of a number of factors: ease-of-use; speed of response; meeting of the requirements of diverse user groups, etc.. It should be noted that user-centred evaluation will have an exploratory nature in the context of the project, mainly for two reasons: (i) MultiMatch search services will be prototypes in the context of the project duration, working on limited amounts of data; and (ii) evaluation of the second prototype has to be carried on in less than three months. Consequently, usability testing, interviews with users and qualitative feedback will have a more primary role in the evaluation than comparative field trials using example user queries and monitoring user search behaviour.

¹ See <http://www.clef-campaign.org/>