Mixed-Initiative Translation of Web Pages

Michael Fleming and Robin Cohen

Department of Computer Science,
University of Waterloo,
Waterloo, Ontario, Canada N2L 3G1
{mwflemin,rcohen}@uwaterloo.ca

Abstract. A mixed-initiative system is one which allows more interactivity between the system and user, as the system is reasoning. We present some observations on the task of translating Web pages for users and suggest that a more interactive approach to this problem may be desirable. The aim is to interact with the user who is requesting the translation and the challenge is to determine the circumstances under which the user should be able to take the initiative to direct the processing or the system should be able to take the initiative to solicit further input from the user. In fact, we envision a need to support interactive translation of Web pages as the World Wide Web becomes more accessible to people with varying needs and abilities throughout the world.

1 Overview

Our main research interest is in the development of guidelines for the designers of mixed-initiative artificial intelligence systems. Mixed-initiative systems are ones where the system can take the initiative to solicit further input from the user or where the user can take the initiative to interrupt the processing of the system to provide additional input (see [11]). Mixed-initiative systems contrast with other AI systems which perform automated reasoning given input from a user, producing output without further interactivity.

To date, the design of mixed-initiative AI systems has been largely done independently by researchers in various subareas of AI, including intelligent tutoring systems ([1, 14]), planning systems ([2, 6, 7]) and interface agents ([6, 8, 13]). Our aim is to determine some guiding influences to direct the design of mixed-initiative AI systems in a more principled manner. To this end, our plan is to study the requirements of various kinds of mixed-initiative applications.

One area which we feel is promising for our study is that of machine translation – in particular, an environment where machine translation is being used to translate web pages for a user. Below we describe our view of this problem and how it should be addressed. We include some comparison to related work and some discussion on how such a system could be evaluated.
2 Translating Web Pages

Our view of the task of translating web pages for users is as follows. First, we feel that users do in fact want translations of web pages – that there is a need for supporting this task. Moreover, we are concerned that the task of translating the entire web page, using a general purpose machine translation (MT) system, is too difficult. We feel that users are probably requesting translations of web pages because they have a specific interest in that web page – there is a particular piece of information being sought within the web page, so that not all of the web page needs to be translated for them. One suggestion is to have a summary of the web page presented to the user first, instead of a full translation – we mean here a translated summary. If a translated summary is desired, one must decide whether to summarize within the language of the text and then translate or to do a rough translation of the entire page and then apply summarization; we hypothesize that summarizing first would be more effective – this is in a scenario where the tool to achieve summarization is more robust that the tool to do translation. In order to produce a summary which maximizes the possible benefit to the user, it would be worthwhile to interact with the user, to determine the reason he is interested in the web page. (Note that this may be true of doing summarization without translation as well).

Some translation systems like Babelfish\(^1\) produce crude translations with some obvious errors. Figure 1 shows some examples of phrases from an original web page and its translation into French, using Babelfish, indicating in the third column various types of errors which were produced in the translation. The question is whether the user can put up with the errors, reading through them to some extent to get the information he requires from the page. In the case of the web page translated in Figure 1, the user requesting the translation had some facility in both languages, so was not adversely affected by the errors. However, in many cases, the user would not know the source language and may be confused by mistakes such as literal translations and incorrect choices of word sense. In any case, there were significant segments of the translation which were acceptable. It is worth noting that the translation system used to produce this output does not model the grammar of the target language\(^2\), hence the cause of some of the errors.

A related question is whether the gist of the translation is really all that a user needs to take away. Resnik\(^1\) in fact claims that getting the gist of a document is often sufficient for a user, who can then decide whether to use other resources to get help with a complete translation. Another question is whether there is more room for interactivity with this kind of a translation process (again, this is the case where perhaps the entire page is translated (or an attempt is made to do so)).

\(^1\) http://babelfish.altavista.com

\(^2\) This information is based on a personal conversation with Elliott Macklovitch at the University of Montreal in 1999.