MSC-Based Formalism for Automated Web Navigation

Vicente Luque Centeno, Carlos Delgado Kloos, Luis Sánchez Fernández, and Norberto Fernández García

Departamento de Ingeniería Telemática
Universidad Carlos III de Madrid,
Avda. Universidad, 30, E-28911 Leganés, Madrid, Spain

Abstract. This article presents an approach to model navigation tasks on the Deep Web [3] with a well known Software Engineering formalism, namely Message Sequence Charts [2] standard from the ITU, in combination with W3C XPath [4] expressions. This modelling can be used to build Wrapper Agents [1] that might automate Web navigation for the user.

1 MSC-Based Formalism

Web navigation may be expressed in terms of MSC [2] components. Both Web clients and Web servers may be represented by **MSC instances** (vertical lines). HTTP requests and answers may be represented by **MSC messages** (horizontal arrows) communicating MSC instances. Vertical axis is also considered as a time axis (top levels are executed before bottom levels).

However, Web navigation not only consists of a single HTTP transaction. Navigation through the Deep Web [3] requires several links to be followed and several forms to be properly filled in. This requires considering the document's internal structure as well. W3C XPath expressions can be used to properly choose which links should be followed next or how should forms be filled in. **MSC actions** (rectangles) might be used for embedding XPath-based data extraction rules as well as other user-defined routines. Figure 1 shows two instances: a Web client and a Web server. After the Web client executes some procedure `A()`, it submits a filled-in form to the server by an HTTP POST request. Once the Web server receives that request, the corresponding procedure `B()` is executed (maybe a CGI program or a servlet) to handle that request and an answer page is returned back to the client. Only when the Web client receives the answer message, it starts executing the `C()` procedure.

Some decisions have to be made during navigation, just as a user behind a browser would do. For instance, it is common that a Web link has to be followed only if some condition occurs, perhaps following other link otherwise. **MSC's inline expressions** may represent these alternative and repetitive behaviours. Just as repetitive or alternative navigation behaviours may be internally structured, also MSC inline expressions may nest.

N. Koch, P. Fraternali, and M. Wirsing (Eds.): ICWE 2004, LNCS 3140, pp. 591–592, 2004.

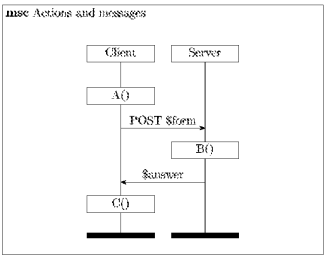

Fig. 1. Instances, messages and actions

Modularized components or parameterized navigation procedures may also be expressed with **MSC references** (curved corner rectangles). Figure 2 shows both an example of a reference to a sub MSC named Identif and its definition. MSC references allow the definition of complex MSC in terms of smaller parts.

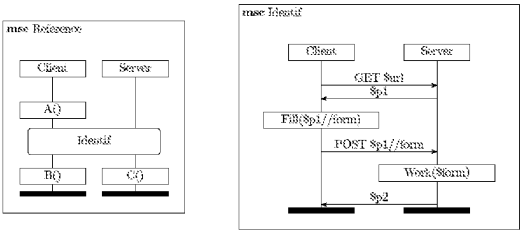

Fig. 2. MSC reference

Acknowledgements. The work reported in this paper has been partially funded by the project Infoflex *TIC2003-07208* of the Spanish Ministry of Science and Research.

References

1. V. L. Centeno, L. S. Fernandez, C. D. Kloos, P. T. Breuer, and F. P. Martin. Building wrapper agents for the deep web. In *Third International Conference on Web Engineering ICWE 2003, Lecture Notes in Computer Science LNCS 2722, Ed. Springer*, pages 58–67, Oviedo, Spain, July 2003.
2. ITU-T. Recommendation z.120: Message sequence chart (msc). In *Formal description techniques (FDT)*, Geneva, Switzerland, 1997.
3. M. P. Singh. Deep web structure. *Internet Computing*, 6(5):4–5, Sept.-Oct. 2002.
4. W3C. Xml path language (xpath) 2.0. *W3C Working Draft 02 May 2003*, 2003.