Prototype of a Discovery Tool for Querying Heterogeneous Services

D. Egret, P. Fernique and F. Genova

CDS, Strasbourg, France, E-mail: egret@astro.u-strasbg.fr

Abstract. The CDS has recently developed a tool for managing URLs in a context of distributed heterogeneous services (GLU, Uniform Link Generator: Fernique et al. 1998). This includes the development of a 'URL dictionary' maintained by the data providers contributing to the system.

Based on such a system, it becomes possible to create automatically a homogeneous interface to the services described in the dictionary. This tool is currently available as a prototype, under the name of AstroGLU, with the aims of demonstrating the feasibility of the concept, and helping orienting future developments in the domain.

Being a flexible, easily maintainable tool, AstroGLU is a strong incentive for an increased cooperation of all astronomical data providers.

1. Introduction

How to help the user find his path through the jungle of information services is a question which has been raised during the past years (see e.g., Egret 1994), when it became clear that a big centralised system was not the efficient way to go.

Obviously the World-Wide Web brought a very interesting medium for solving this question: on the one hand the WWW provides a common language for all information providers (but flexible enough so that it does not place unbearable constraints on existing databases); on the other hand the distributed hypertextual approach opens the way to navigation between services (provided a minimum of coordinating spirit can be achieved). Let us note that it has been already widely demonstrated that coordinating spirit is not out of reach in a (small) community such as astronomy, which also remains largely sheltered from commercial influence.

The CDS (Centre de Données astronomiques de Strasbourg) has recently developed a tool for managing remote links in a context of distributed heterogeneous services (GLU, Générateur de Liens Uniformes, i.e., Uniform Link Generator; Fernique et al. 1998). First developed for ensuring efficient interoperability of the several services existing at CDS (Astronomer's Bazaar, VizieR, SIMBAD, bibliography, documentation, etc. — see Genova et al. 1996), this tool has also been designed for maintaining addresses (URLs) of remote services (ADS, NED, etc.).

Capacity of leaving the management of these addresses to the remote data provider (rather than having to centralize all the knowledge of the distributed services) is an essential feature of the GLU.

A key element of the system is the 'GLU dictionary' maintained by the data providers contributing to the system, and distributed to all machines of a given domain. This dictionary contains knowledge about the participating services (URLs, syntax and semantics of input fields, descriptions, etc.), so that it is possible to generate automatically a correct query to be submitted.

A typical scenario is the following: we want to get from service S, a piece of information J, corresponding to the data D: the GLU system will ensure that the correct query is generated, i.e., in WWW syntax:

$$\texttt{address\_of\_S/query\_for\_J?field\_D}$$

In fact, the remote user has no visibility of the GLU: GLU is essentially a tool for the data providers. The user will simply see a series of features accessible in a sequence, such as, for example, once an author name is provided (by the user, or retrieved from a database in a previous step), it becomes possible to press a button and to obtain the list of all papers published by this author (from ADS), or the list of recent preprints (in astro-ph) mentioning his name. How the button leading to this information is generated remains transparent to the user, who probably does not care to know (even if, hopefully, he or she appreciates the opportunity which is being offered).

The service provider (data center, archive manager, or Webmaster of an astronomical institute) has used the GLU for coding the query, making use of the easy update of the system: knowing which service to call, and which answer to expect from this service, the programmer does not have to worry about the precise address of the remote service at a given time, nor of the detailed syntax of the query (expected format of the equatorial coordinates, etc.).

2. What can we find about ... ?

Let us imagine, now, another scenario: we have the data D (for example an author's name, position or name of an astronomical object, bibliographical reference, etc.), and we would like to know more about it, but we do not know which service S to contact, and what are the different types of information J which can be requested. While the first scenario was typical of an information provider (who knows the astronomical landscape of information services, and has developed contacts with the managers of interesting remote databases), this latter scenario is typical of a scientist, exploring new domains as part of a research procedure.

2.1. A Reference Directory

The GLU dictionary can also be used for helping to solve this question: the dictionary can be considered as a reference directory, storing the knowledge about all services accepting data D as input, for retrieving information types J, or K. For example, we can easily obtain from such a dictionary the list of all services accepting an author's name as input : information which can be

accessed, in return, may be an abstract (service ADS), a preprint (LANL/astro-ph), the author's address (RGO e-mail directory) or his personal Web page (StarHeads), etc.

Based on such a system, it becomes possible to create automatically a simple interface guiding the user towards any of the services described in the dictionary.

2.2. AstroGLU

This idea has been developed as a prototype tool, under the name of AstroGLU[1], in order to demonstrate the feasibility of the concept, convince more data providers to collaborate, and help orienting future developments in the domain.

The current steps of a usage scenario are the following:

1. Data Type Selection: First, the user can select among the available data types those which correspond to the data D for which additional information is needed. The principal data types already available in the current prototype version (3.0, September 1997) are the following: name of an astronomical object, celestial position, last name of a person (e.g., an author's name), keyword (in English natural language), reference code (bibcode used by NED/SIMBAD/ADS), catalog name, dataset number in HST or CFHT archive, etc.

At this stage, the user can already input the data D itself, assuming it is a simple character string (e.g., an astronomer's name). But the user does not need to have an *a priori* knowledge of the existing services, or even of the information types corresponding to potential query results.

2. Service List: Based on this data type, AstroGLU scans the dictionary and selects all services (among those known from the system) that support queries involving this data type; for example, if the data type is 'astronomical object name', SIMBAD, NED, or the ALADIN sky atlas (among others) are listed.

This list is supposed to answer simultaneously two questions: *what ?* (i.e. which type of information can be found) and *where ?* (i.e. which service can provide it). But the focus is made on the first aspect, the second one being generally kept implicit until the selection of the information type is done.

3. Query Submission: The user can finally select one of the proposed services, and will receive a form for submitting the query to the remote service, to which it is finally submitted for processing. These forms frequently imply giving additional parameters in complement to the data D (e.g., epoch of a position; year limits for a bibliographical query, etc.).

Where can we find ... ?   Alternatively, the user can specify what he "looks for", and according to the service qualifications contained in the dictionary, the user will be presented with a selection of services able to answer his query.

---

[1]http://simbad.u-strasbg.fr/demo/cgi-bin/astroglu-m1.pl

3.   Current AstroGLU Functionality

AstroGLU functionality is constructed around the main dictionary. In step 1, all data types listed are those occuring at least once in the dictionary (or, more specifically in the subset of the dictionary related to the specific domain on which AstroGLU is working). These data types may be sorted according to eventual conversions (using remote resolvers, or local rules). At the end, they are sorted in alphabetical order.

In step 2, the list of 'actions' (i.e. possible queries) using the data type selected in step 1 is displayed, and, if a data string has been given, some tests can be performed, when a test method has been implemented in the dictionary (e.g., compliance of a refcode with the corresponding dataset). Some of these actions may imply use of one, or more, intermediate resolution (e.g., call SIMBAD for finding the celestial position of a given object, before sending a query to an archive).

Step 3 may include, in the future, examples and default values of additional parameters.

The complete list of actions can be displayed on request.

AstroGLU can be automatically implemented for all or part of the domains cooperating to the GLU system. All the forms are generated from the GLU Dictionary information. The dictionary being very easily and efficiently maintained, this is a strong incentive for an increased cooperation of astronomical data providers.

4.   Final Remarks

A major aim of this tool is to help the user find his way among several dozens (for the moment) of possible actions or services. A number of compromises have to be made between providing the user with the full information (which would be too abundant and thus unusable), and preparing digest lists (which imply hiding a number of key pieces of auxiliary information, and making subjective choices).

A resulting issue is the fact that the system puts on the same line services which have very different quantitative or qualitative characteristics. Heck (1997) has frequently advocated that high quality databases should be given preference, with respect to poorly updated or documented datasets. We do not provide, with AstroGLU, efficient ways to provide the user with a hierarchy of services, as a gastronomic guide would do for restaurants. This might come to be a necessity in the future, as more and more services become (and remain) available.

References

Egret, D. 1994, in ASP Conf. Ser., Vol. 61, Astronomical Data Analysis Software and Systems III, ed. Dennis R. Crabtree, R. J. Hanisch & Jeannette Barnes (San Francisco: ASP), 14

Fernique, P. et al. 1998, this volume

Genova, F. et al. 1996, Vistas in Astronomy 40, 429

Heck, A. 1997, in "Electronic Publishing for Physics and Astronomy", Astrophys. Space Science 247, Kluwer, Dordrecht, 1 (ISBN 0-7923-4820-6)