

Collection of Web tables

Part of 181.130 VU [Web Data Extraction and Integration](#)

Date: 19.10.2006

Deadline: Thursday, **26.10.2006**

Contact person in case of questions: Wolfgang Gatterbauer (gatter @ dbai.tuwien.ac.at)

Background

Some ongoing research at DBAI focuses on information extraction from web tables. In order to objectively test and evaluate different algorithms, we are looking for a collection of different web pages that contain web tables with information relevant to different internet users. Especially, we want to know about web tables that are of interest to people who are currently not involved in the project, and as such not aware of the methods we use. That's why we are interested in web tables of interest to you.

Your task

- (1) Gather the URLs of **5 web pages that contain different web tables** of certain interest to you. "Of certain interest to you" is defined very broadly. Basically, the tables should contain some kind of information that you might be happy to obtain in an automated fashion.
- (2) Write an email to Wolfgang Gatterbauer (gatter @ dbai.tuwien.ac.at) containing the URLs of those 5 web pages. Use "VU Example tables" as subject.

Your chosen 5 web tables should look different from each other, which means they should not be built upon the same template and should cover different subject areas. Most likely, they will, therefore, be from different websites.

Your 5 web pages have to be accessible just by their URLs and should not need additional navigation from one entry page to the page containing the table of interest.

What do we do with these tables

We will use your tables to perform some tests and to evaluate our system. The results of these experiments together with the algorithms we use will be explained and discussed as part of the lecture, either on December 12th or January 11th.