Sounds great? You want to know more? Please contact us!

www.isvworld.com
1 Overall Objective
Provide market insight and a collaboration platform for & by European Independent Software Vendors, the Scientific World and Governmental Organizations to support the growth of the European Software Industry.

With the term Independent Software Vendor we refer to any company developing software products.

- ISVWorld delivers ISV-related market insight necessary for the customers to make the right decisions, every day.
- A "Film" rather than a "Photo" of the European Software Industry.
- Create and share insight into the European Software Industry community, monitor trends and evolution.
- Become the preferred unified European source of choice to identify ISVs, software products and market insights.
- Using a “Web 2.0” community approach, make it a platform for & by European ISVs, Scientific World and Governmental Organizations
- Bring European ISVs across borders together, and closer to its customers.
- A automatically aggregated database of ISVs that is bigger/more complete than any other database, thanks to among others innovative / intelligent use of taxonomy-driven information retrieval technology.

2 Candidate
We are looking for an information technology student (bachelor or master) who can help developing a multithreaded application which automatically gathers URLs, using free crawling techniques, extracts their textual content, and identifies their language (for more information on the available projects, see chapter 5). The application will be based on technologies like Java, Python and MySQL on a linux platform. We expect the student to be familiar with technologies like object oriented programming (OOP), design patterns, UML, and the basics of Java, multithreaded development, SQL and linux development. A ‘stagevergoeding’ is on offer from us and we are located in Utrecht. The project can start at any moment.

3 What can we offer you?
ISVWorld is, as a innovative start-up, a unique and rewarding place to work, whether you are just starting out or wish to advance and develop your career. We will actively encourage you to acquire skills and knowledge, which will add to your experience. This benefits both you, and ISVWorld.

Sounds great? You want to know more? Please contact us!
Among others, you will be able to apply the key concepts of multi-threading and information retrieval technologies. Salary has to be defined. Moreover, you will be working in a team of seven people, with many talents. Among them is René van Erk, who is member of the European management team of Wolters Kluwer where he is responsible for all Product - and Business Development. In this role, his key responsibility is to optimize the WK portfolio for maximum growth.

4 Sounds great? You want to know more?
You can contact Ivo Hunink, program manager of ISVWorld. You can reach him at ivo@isvworld.com and 06 23 54 63 08. We are looking forward to hear from you!

5 Projects
One of the main components of ISVWorld is the automatically aggregated database of ISVs. Innovative / intelligent information retrieval technology is used to gather the necessary information for this database. The development of this technology is divided in several sub-projects of which two can be developed by an enthusiastic student.

5.1 Sub-project 1
The goal of this project is:

“To develop an application which automatically gathers URLs, using free crawling, extracts their content, and identifies their language”

This application consists of several main components:

1. A component that is able to extract URLs from:
   a. Directory Sites (like LinkedIn, softwarepakketten.nl, etcetera). Preferable from only those parts of the directory sites which are targeted to ISVs, for example [http://www.linkedin.com/directory/companies/computer-software-engineering.html](http://www.linkedin.com/directory/companies/computer-software-engineering.html).
   b. Using the Bing search engine API and/or Google search engine API, using queries like “Software company in Groningen”, “Accountancy Software in Utrecht”
   c. Free crawl. Just starting at one point, and just following any link available.
2. A component that fetches all the content of the URLs gathered at component 1.
3. A component that is able to identify the language of the content, preferably using existing open source tools.
4. A component that is able to translate non-English content to English (using for example the Google translate API).

Sounds great? You want to know more? Please contact us!
The end result will be a database containing URLs, their source, their language, and their content. In the end, this content can be used to identify whether or not that specific URL is a real website of an ISV (see project 2).

5.2 Sub-project 2
The goal of this project is:

“To develop an application which automatically identify an ISV’s website using the data collected at project 1”

This application consists of several main components:

1. A component that is able identify whether or not a website belongs to a company.
2. A component that is able to identify whether or not a belongs to an ISV.

To test the components, we have testing data available which is collected through manual labor. In the end, for each URL, this application will identify the score in % that a website is an ISV or a non-ISV website.