

Tapaus 2

Olet töissä suuressa lehtitalossa. Tehtävänäsi on suunnitella hakurobotti, joka käy läpi kotisivuja ja poimii sieltä potentiaalisten tilaajien postiosoitteita ja puhelinnumeroja. Miten määrittelet, millaiset merkkijonot sisältävät kelvollisen osoitteen?

Voit esimerkiksi tutkia Kissastanian Logiikkakoulun henkilökunnan ja opiskelijoiden kotisivuja:

<http://www.cs.joensuu.fi/pages/whamalai/tepe04/kissastania/kissastania.htm>.

Miten löytäisit niiden kissojen henkilötiedot, joille kannattaisi lähettää näytenumero uudesta lehdestä "Scientific MIU-Systems"?

You working in a journal publishing company. Your task is to design a search engine, which scans through the home pages and looks for postal addresses and telephone numbers of potential subscribers. How do you define, what kind of pages contain a proper mail address?

As an example, you can look at Kissastania Logic School staff and students' homepages:

<http://www.cs.joensuu.fi/pages/whamalai/tepe04/kissastania/kissastania.htm>.

How could you find personal data of those cats, who might be interested in a sample copy of a new journal "Scientific MIU-Systems"?