Large-scale community protein annotation -- WikiProteins

May 28 2008

Today sees the launch of a new collaborative website initially focusing on proteins and their role in biology and medicine. The WikiProfessional technology underlying the site has been developed based upon the collaborative Wikipedia approach. Described in BioMed Central's open access journal *Genome Biology*, WikiProteins provides a method for community annotation on a huge scale.

The article is written by Barend Mons of the Erasmus Medical Center in Rotterdam, and the Leiden University Medical Center, The Netherlands, and his co-authors from Brazil, The Netherlands, Switzerland, the UK and the USA. They include Amos Bairoch of UniProt, Michael Ashburner of GO and Jimmy Wales, the co-founder of Wikipedia.

The source material for WikiProteins comes from a mixture of existing authoritative databases (such as the Unified Medical Language System, UniProtKB/Swiss-Prot, IntAct and GO), supplemented by concepts mined from scientific papers published in public literature databases. The automated data mining identifies 'facts' in these available resources, such as protein functions or protein-disease relationships. This process created over one million biomedical concept clouds – called 'Knowlets' – around each individual concept.

The developers of the site now hope that many researchers will follow their call to annotate, via WikiProteins, the Knowlets for which they are leading experts. The method enables researchers to add data even from sources that are not openly available, such as from journals only

accessible via publishers' databases, immensely enhancing the potential for comprehensive coverage. Each page of text called up via the system is automatically indexed and concepts are connected to the WikiSpace, so that their definition comes up and the information can be edited directly from the page.

The resulting data in the Wiki is fully and freely accessible to the public, and entries can be annotated by any registered user. Mons said: "We here call on a million minds to annotate a million concepts and collect new facts from full-text literature with the immediate reward of collaborative knowledge discovery and recognition of Wiki-contributions to the scientific community."

A preview of the WikiProtein technology is available at: conceptweblinker.wikiprofessio … .org/monsarticle.htm

Source: BioMed Central