

TIGERSearch attacks Proteins

Uwe Reyle ⁽¹⁾, Jasmin Saric ⁽²⁾, Wolfgang Lezius ⁽¹⁾, Esther Koenig-Baumer ⁽¹⁾

GenIE Genome Information Extraction

(1) Institut für Maschinelle Sprachverarbeitung, IMS, University of Stuttgart, Germany

(2) European Media Laboratory (EML), Heidelberg, Germany

It is common usage to associate particular features to regions and other sites of interest in the sequence of proteins. Based on an ontological model of proteins, protein structure and functions of (parts of) proteins we automatically translate the content of Swissprot feature tables into fully structured representations that may be efficiently queried by the TIGERSearch engine.

TIGERSearch offers search on tree-like feature graphs that – in addition to the usual part-whole relations between amino-acid residues and any regions of interest of a protein -- accommodate the representation of the ternary structure of a protein (any kind of bond between two peptides is represented by an edge) as well as the modelling of non-contiguous domains. Furthermore with each node of the graph a feature record may be associated that contains further and in particular non-structural information (like posttranslational modifications, binding sites, enzyme active sites, etc.) attributed to the amino acid or region in the Swissprot feature table.

A major reason to use TIGERSearch for search on protein structure is provided by its query language. It offers the possibility to search for structural features encoded by the graph and non-structural features as given in the feature records at the same time. The use of variables and regular expressions makes the query language very powerful, elegant and easy to learn.