

## TaxonConcept - TaxonRank a synonymy ranking algorithm for earth science data networks

Enough species have been described and investigated to make them extremely important tools in e.g. stratigraphy or paleoclimatology. Such large volumes of information however are only useful if effectively organized and made accessible, so that needed data can be quickly identified and retrieved. The classic method for doing this in paleontology is publishing primary taxonomic data in journals by carefully prepared synonymy lists. Most journals demand the use of the so-called 'Open Nomenclature', which allows working with taxonomic classifications that are unclear and allows the author to comment the identification of specimen of other authors.

These synonymy lists rather reflect the taxonomic opinion of its author than providing objective synonymies. Therefore, the evaluation of the quality of these taxonomic identifications are very useful for the application of e.g. biostratigraphy by non-paleontologists. But for a geoscientist not specialised in the taxonomy of the species in question it is difficult to follow the existing taxonomic literature and to find those taxa for which related data may exist and which possibly have been treated previously as synonyms of other taxa.

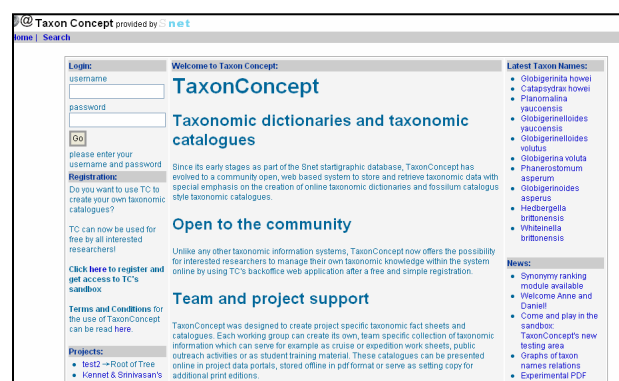
Essentially, a synonymy list is a list of citations related to a taxon name, annotated in a specific way to express an author's opinion on these synonymies. The bibliographic nature of a synonym lists should lend itself to bibliometric techniques to investigate the ranking of synonymy entries and used taxa, respectively. We present TaxonRank, a new ranking algorithm based on bibliometric analysis and internet page ranking technologies.

TaxonRank uses published synonymy list data stored in TaxonConcept, a taxonomic information system which is part of the Stratigraphy.Net project.

The bibliometric algorithm is then modified by the certainty of a species identification based on the open nomenclature notation used in the synonymy list, as well as other synonymy specific criteria which will be explained.

Synonymies are a well known a problem for earth science databases which mostly leave the original classification by the dataset's author unchanged. Some databases are therefore difficult to query and might encounter difficulties when they try to build data networks. Technologies like TaxonRank will help those networks to create sophisticated ontologies which enable the data user to formulate adequate queries.

Internet: <http://www.taxonconcept.stratigraphy.net/>



The screenshot shows the TaxonConcept website interface. At the top, it says "TaxonConcept provided by n e t". Below this, there is a navigation bar with "Home" and "Search". The main content area is divided into several sections:

- Login:** A form with fields for "username" and "password", and a "Go" button.
- Registration:** A section asking "Do you want to use TC to create your own taxonomic catalogues?" and "TC can now be used for free by all interested researchers!". It includes a link to "Click here to register and get access to TC's sandbox".
- Terms and Conditions:** A link to "Terms and Conditions for the use of TaxonConcept can be read here".
- Projects:** A list of projects including "test2 - Root of Tree" and "Kornet & Simeone's (1983) Neogene".
- Welcome to TaxonConcept:** A central section with the heading "TaxonConcept" and "Taxonomic dictionaries and taxonomic catalogues". It contains introductory text about the system's evolution and a link to "Open to the community".
- Open to the community:** A section stating "Unlike any other taxonomic information systems, TaxonConcept now offers the possibility for interested researchers to manage their own taxonomic knowledge within the system online by using TC's backoffice web application after a free and simple registration."
- Team and project support:** A section explaining that TaxonConcept was designed to create project-specific taxonomic fact sheets and catalogues, and provides a link to "TaxonConcept's new testing area".
- Latest Taxon Names:** A list of recent taxonomic entries such as "Globigerinita howei", "Calappastrax howei", "Planorbulina yaucouensis", "Globigerinitoides yaucoensis", "Globigerinitoides volutula", "Globigerinita voluta", "Phanerostomum asperum", "Globigerinitoides asperus", "Heterbergella brittonensis", and "Whiteoensis brittonensis".
- News:** A section with news items like "Synonymy ranking module available", "Welcome Anne and Daniel", and "Come and play in the sandbox: TaxonConcept's new testing area".