PalyWeb: A palynomorph database project on the web

[PalyWeb : un projet de banque de données de palynomorphes sur le web]

Philippe STEEMANS[1]

Pierre BREUER[2]

Key Words: Palynology; database; Internet

Citation: STEEMANS P. & BREUER P. (2007).- PalyWeb: A palynomorph database project on the web. *In*: STEEMANS P. & JAVAUX E. (eds.), Recent Advances in Palynology.- Carnets de Géologie / Notebooks on Geology, Brest, Memoir 2007/01, Abstract 08 (**CG2007_M01/08**)

Mots-Clefs : Palynologie ; banque de données ; Internet

Introduction

Since the sixties, palaeopalynology has been experiencing a spectacular growth. Already by the end of 1966, KREMP (*in* LENTIN *et alii*, 1996) had calculated that about 330 new articles on palynology were published each year in more than 200 journals, and about 4200 had already been published. In a paper published during the 90's, JANSONIUS and MCGREGOR (1996) made an estimation of the number of palynological papers published each year. Although they noted that data are incomplete, around 400 new articles are published annually. Because of this, palynology is a victim of the abundance of published information, which has created taxonomic chaos in some palynological groups. In addition, numerous palynomorphs have been published invalidly, some with redundant names. Several taxa have been described in publications with a limited distribution in other countries or in languages not accessible to most palynologists. Therefore, palynologists were soon convinced that an urgent need existed to create catalogues, or large datasets to manage this mountain of data. The first of these were in a paper format but it rapidly became clear that the solution to the management of palynological data is the use of personal computers. The better known applications for this purpose, like PALYNODATA, DinoSys, etc. are briefly described in LENTIN *et alii* (1996), but there are other projects: for the Palaeozoic, Phytopal for acritarchs and Chitinovosp for Chitinozoans.

The MediaWiki technology

Originally written for the well-known Wikipedia encyclopaedia, MediaWiki is free software accessible on the Internet (http://en.wikipedia.org/wiki/), but is now run on other projects of the non-profit Wikimedia Foundation (http://en.wikipedia.org/wiki/Wikimedia_Foundation/) and many other wikis. The term wiki is a short form of "wiki wiki" which means something quick or fast in Hawaiian. On the Internet, a wiki is a website that allows users to add, remove, or edit content very quickly and easily.

PalyWeb is a new palynomorph database project based on MediaWiki technology. It is a web-based open-access, free-content palynological database. To our knowledge, it is alone in its mode of data acquisition and management for palynology. It is also the only one to have as an objective the presentation in one dataset of all fossilised palynomorph taxa. Its conception is similar to that of the Wikipedia encyclopaedia, so it can be considered a wiki website. The ease of interaction and operation makes a wiki website an effective tool for collaborative scientific writing. **PalyWeb** is designed to be a database compiled collaboratively by volunteers, and allows most articles to be changed by anyone with access to a computer, a web browser and an Internet connection. However, users must be logged in to avoid vandalism and inconsistency. **PalyWeb** is built with the expectation that over time collaboration among users will improve the quality of the articles.

[1] *Laboratoire de Paléobotanique, Paléopalynologie et Micropaléontologie, Université de Liège, Allée du 6 août, B18, Sart-Tilman, 4000 Liège (Belgium)*
p.steemans@ulg.ac.be
[2] *Laboratoire de Paléobotanique, Paléopalynologie et Micropaléontologie, Université de Liège, Allée du 6 août, B18, Sart-Tilman, 4000 Liège (Belgium)*
piet79@yahoo.fr
Manuscript online since March 22, 2007

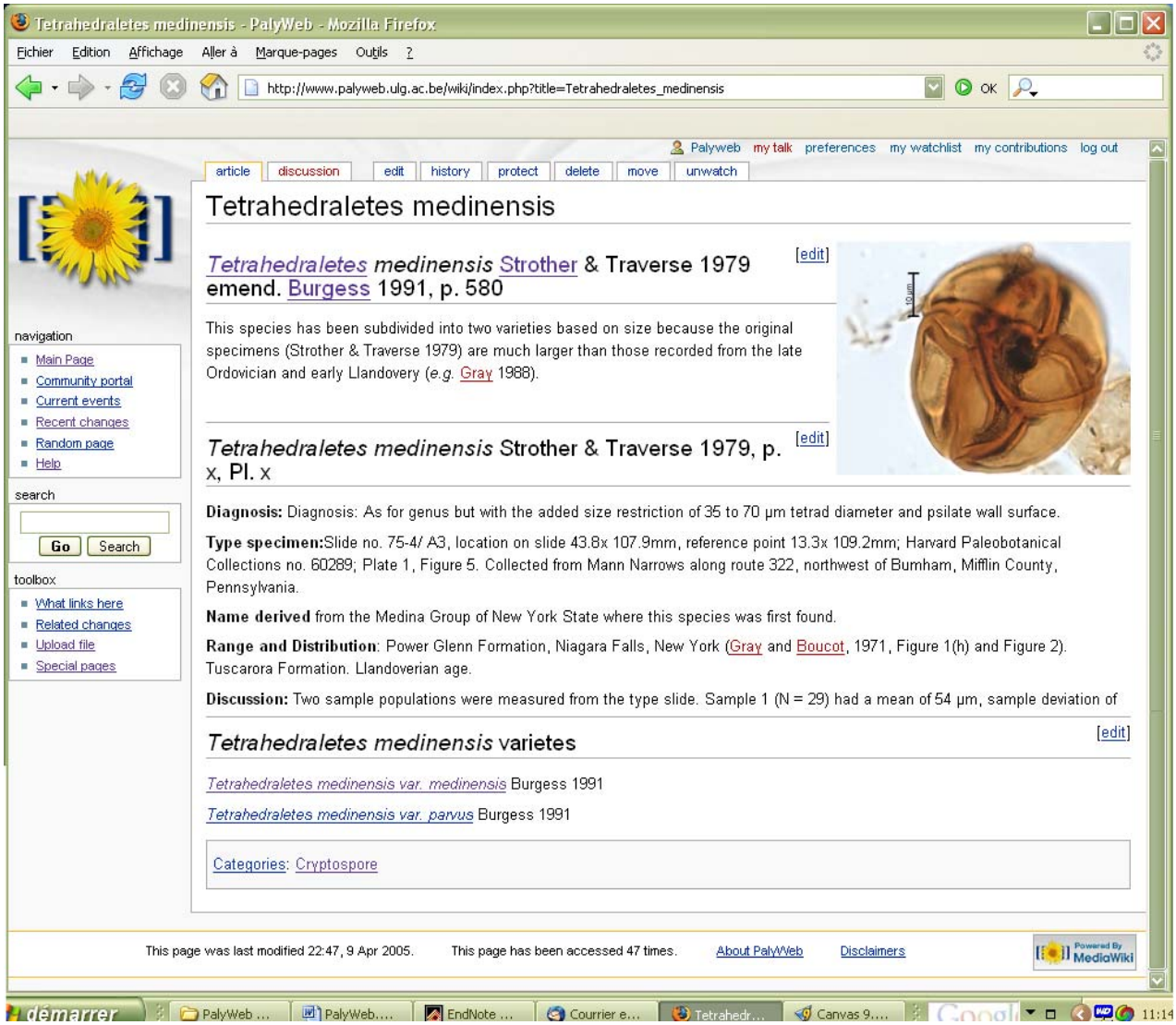

Figure 1: Example of a **PalyWeb** page containing the description of the *Tetrahedraletes* genus. Bleu words inside of the descriptive text are links to other pages (bibliographic references or genus description pages).

The **PalyWeb** project is housed on the web server of the University of Liège. A daily backup is carried out to safeguard all information in the event of failure of the system. Its Internet address (URL) is: http://www.palyweb.ulg.ac.be/wiki/. To give an estimation of the volume of information that the operating system can manage, the Wikipedia website provides some statistics: on May 2006, Wikipedia had 1,122,525 articles. That number excludes redirects, discussion pages, image description pages, user profile pages, templates, help pages, portals, articles without links to other articles, and pages about Wikipedia. Including these, Wikipedia has 4,154,971 pages. Users have made 53,926,940 edits, an average of 12.98 per page, since July 2002. At that time Wikipedia had 1,395,387 registered user accounts.

The PalyWeb project

PalyWeb is designed to accept the descriptions of a maximum number of published palynomorph taxa available from conventional scientific reviews in accordance with the Botanical and Zoological Nomenclature Codes. **PalyWeb** has been deliberately limited to pre-Mesozoic palynomorphs, the field of research of the authors. It was presented to the scientific community officially at the General meeting in Prague (2006) of the "Commission International de la Microflore du Paléozoïque" (http://www.cimp.ulg.ac.be/). **PalyWeb** is not an online medium for publication. New taxa cannot be presented on its website. However, concerning a published species there is no limitation in the number of pages or their length. Text, figures and pictures may be up- and downloaded. Of course, to avoid long delays, pictures may be compressed. It is forbidden to up-load pictures protected by copyright. Navigation through pages may be done by clicking on internal links or by typing keywords in the search engine (*e.g.* the name of species, *etc.*).

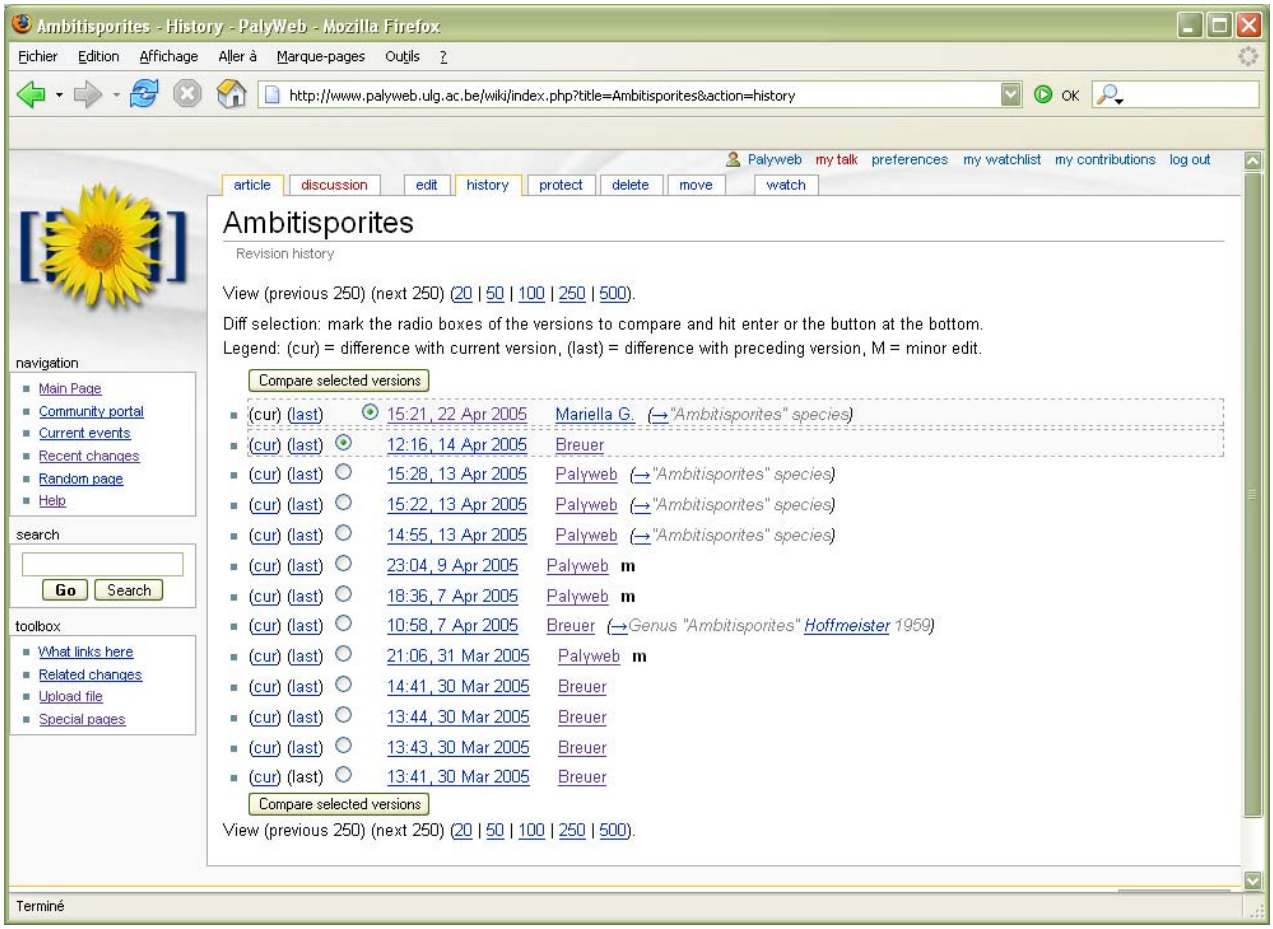

Figure 2: Example of an historic page with the date and the author of the modifications.

At the moment, the database is organized in three sections: the main one hosts the descriptions of taxa previously published in scientific reviews, the second concerns unpublished taxa left in open nomenclature for which information from the **PalyWeb** community is requested, and the third is devoted to enigmatic palynomorphs. The second and third are actually discussion forums, not a part of the database. Of course, this structure is tentative, and can be debated and easily modified. Such flexibility is one of the advantages of a wiki-type website. Each page (article) is articulated around four main components. The first is the "page" itself (Fig. 1) with a title, a name at the top of the screen. The second component is obtained by selecting the heading "to modify" (edit page), where upon the page can be changed immediately, without restrictions. A third component is the heading "discussion", where **PalyWeb** users can exchange their ideas on the contents of the pages. The last component is the "history" of pages (Fig. 2). All former versions of the pages are stored with the user names, the authors of its modifications. Each previous version can be restored (a very useful tool as this procedure prevents mistakes and allows the evolution of the page content to be followed in accordance with modifications made by various contributors).

The general architecture of the website is illustrated on Figure 3 using as an example the organization of a page concerning the cryptospore genus *Tetrahedraletes*. The words in blue are internal links to existing pages. The words are in red when a link has been created but the page to be linked does not yet exist. All words useful for indicating relationships within the database can be converted into links to other **PalyWeb** pages (external links are also possible). Most of the connexions in the example link to taxonomic pages, but is possible to link to pages of other subjects, for example, pages with bibliographic references, personal pages, *etc.* Each page can be classed in several categories. This is a very important tool, not only as an aid to navigation through the website, but also to organize pages in relation to several discrete criteria. It is particularly interesting to group taxa by morphological characteristics, by stratigraphic levels (Fig. 4), by geographic distribution and

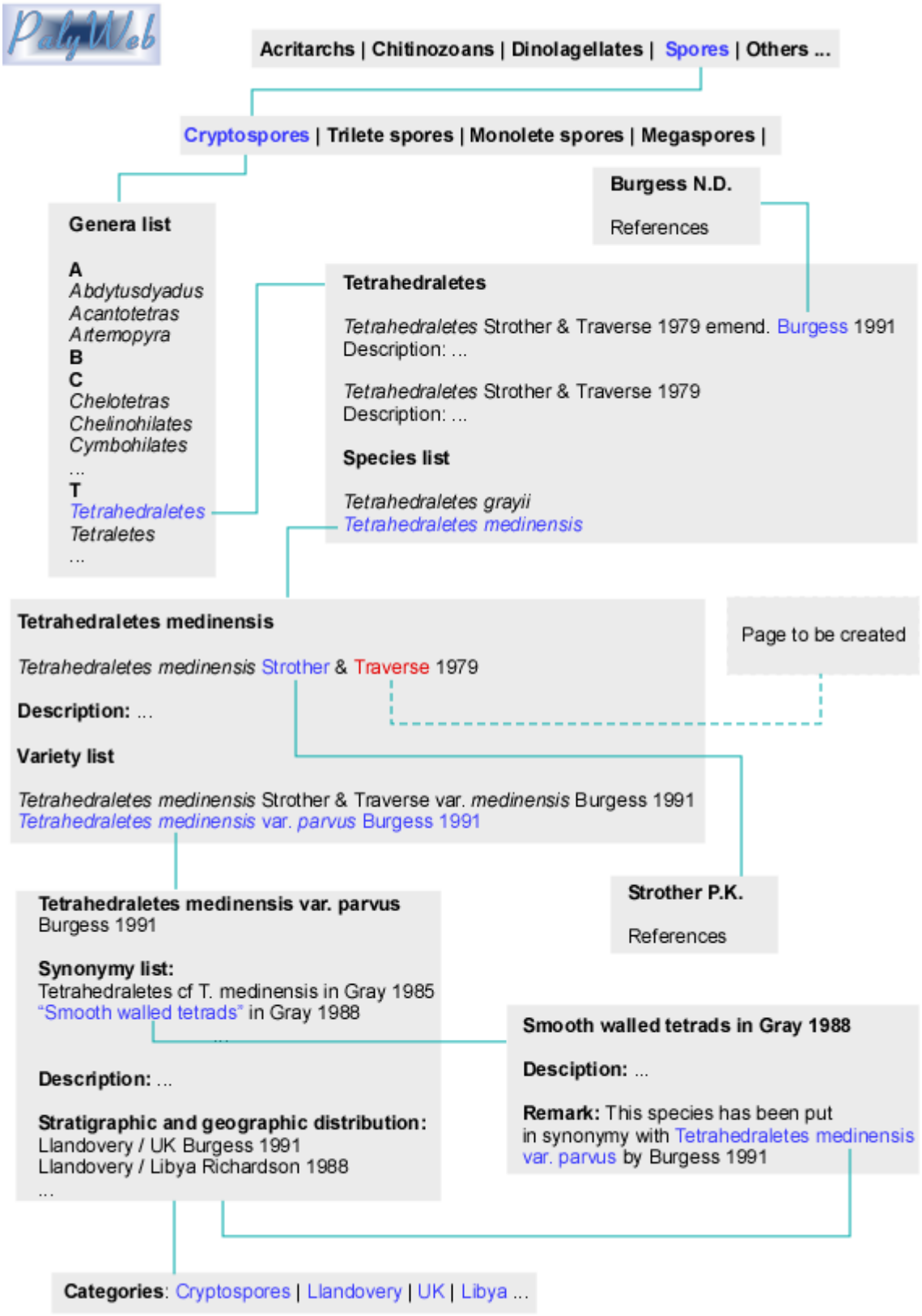

Figure 3: General architecture of **PalyWeb** based on the example of the *Tetrahedraletes* cryptospore genus.

so on. Categories make this possible for the application groups pages automatically and alphabetically by the several criteria. One page can be assigned to several different categories. Categories can be created or modified like a conventional page. However, the great flexibility of a database of this kind could generate problems. To avoid inconsistency, it is important that syntactical rules be established for naming the internal links. For example, in **PalyWeb**, "cryptospore" and "cryptospores" are discrete pages. Therefore, it is necessary to be strict in the way one contributes to the database.

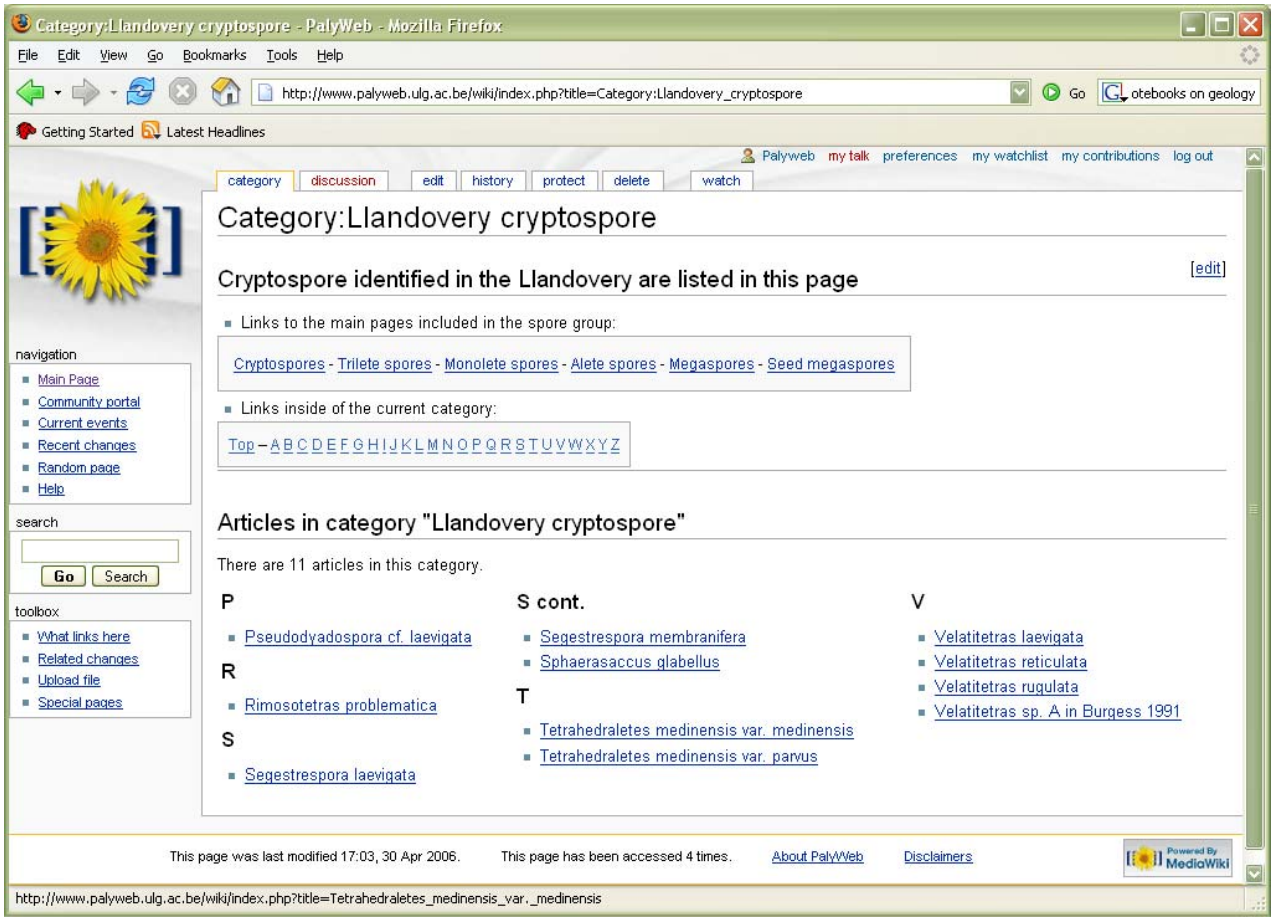

Figure 4: Example of a category page based on the example of Llandovery cryptospores already introduced in the **PalyWeb** database.

The advantages of PalyWeb

The advantages of such a database are numerous. The whole palynological community has free access to the website and everybody is allowed to improve or to add to its contents. The free distribution, constant updates, diverse and detailed coverage, by numerous professional palynologists guaranties the high quality of data. The database is accessible by any computer with an Internet connexion (Windows, Mac, Linux). Common web browsers are adequate for work with the database (Internet Explorer, Firefox, *etc.*). Therefore, it is not necessary to buy new software and to keep it up to date. The database is accessible from wherever you are. To work with a colleague in a foreign country, a copy of the database in your computer is not needed and there is no problem of cross-platform incompatibility. Information introduced into the database is immediately available to all. Unlike other databases, it is not necessary to wait until the database is considered complete before it becomes accessible. In addition, as new palynological data are available continually, databases may be constantly and immediately upgraded. Wait for the new version of a database (usually obsolete in a short time) is eliminated. There is no limitation on size or contents (except the capacity of the disk on which the information is stored). The content of the database is highly flexible and it was concepted to facilitate discussion. The website was not conceived and built by a pugnacious individual or a small group; it is the result of the worldwide collaboration of the whole palynological community. Meetings for the exchange of information are no longer necessary, and in any event, were commonly restricted in number to those with travel subsidies.

References

JANSONIUS J. & MCGREGOR D.C. (1996).- Chapter 1. Introduction. *In*: JANSONIUS J. & MCGREGOR D.C. (eds.), Palynology: Principles and applications.- *American Association of Stratigraphic Palynologists Foundation*, Salt Lake City, n° 1, p. 1-10.

LENTIN J.K., DAVIS O.K., MUNCEY T.S. & PIEL K.M. (1996).- Chapter 24. Personal computers in palynology. *In*: JANSONIUS J. & MCGREGOR D.C. (Eds.), Palynology: Principles and applications.- *American Association of Stratigraphic Palynologists Foundation* Salt Lake City, n° 1, p. 961-984.