ABSTRACT:

THE THESAURUS AS SMALL-WORLD.
THE NET OF OPTIMAL ACCESS OF THERMS

Rudolf Ungváry Rudolf, ungvary@hungary.com

National Széchényi Library

Strong connections in Thesauri are formed by the Guarini terminology based rigid generic relations. If the rigid relations are not complete the Thesaurus falls apart to several different parts, consequently the Thesaurus will not be compact. In case of information retrieval languages, generic relations ensure the strongly-connected nature that has already been described in social networks.

On the other hand, non-rigid relations which play semantic or tematic roles (Parsons, Jackendoff) do not require completeness. In their case it is sufficient that at least a few partitive, causal and other relations are connected to one node. If these weak connections are broken, the compactness of Thesauri do not cease, they do not fall apart. The weak relations ensure that in any given Theasaurus relatively few steps — maximum 5–7 — lead from one node (lexical unit) of the Tesaurus to another. This way the information retrieval language can be characterized by the „small-world" phenomenon (Barabási). (In such worlds everyone knows everyone within 5-6 steps. This term comes from the feeling that "it is a small world" when we meet someone new who knows another person we know). In these worlds the notion of an average of six degrees of separation between the nodes is applicable.

A connectivity of such degree makes possible the stability of the system and optimal navigation inside the system when selecting the most appropriate phrase. Consequently the structure of information retrieval languages is similar to that of (hub-based) scale-free networks.

The study of the above mentioned characteristics can be completed ideally on Thesauri of the size of the broad Thesaurus of the National Library of Hungary, because the number of its lexical units already allows modelling of arbitrary linking steps and consequently that of interconnectivity.