Analysis Methods for the Terminological Structure of a Subject Area

D. A. Gubanov, A. V. Makarenko, and D. A. Novikov

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

Received January 28, 2013

Abstract—This paper suggests an automated expert approach to analysis and synthesis of the terminological structure of a subject area. A key distinction is that the designed procedure employs formal analysis operations to obtain numerical characteristics for the terminological structure of a studied subject area. The basic capabilities of the developed approach are illustrated by the subject area of general methodology.

DOI: 10.1134/S0005117914120133

1. INTRODUCTION

According to the definition stated in [20], a theory is the form of authentic scientific knowledge about a certain set of objects, representing the system of interconnected assertions and proofs and containing the methods of explanation and prediction of phenomena and processes in a given problem domain (subject area), i.e., of all phenomena and processes described by this theory.

Generally, scientific results in a certain subject area are presented on an appropriate professional language [19] which operates, in addition to common lexis, a special terminology.

In other words, for each subject area, one can formally assign a set of terms characterizing this subject area and being used there (we will describe such a procedure below). The set of these terms and their relations (see the discussion below) is called the terminological structure of a subject area.

Let us make a couple of remarks. First, in contrast to theory of terminological systems (e.g., see [1, 2, 10, 11, 27, 31]), we are not concerned with general qualitative regularities of formation and further development of terminology in some field of knowledge. Instead, our intention consists in studying the quantitative properties of its current “slice,” e.g., by analogy to the paper [25] focused on the subject area of network-centrism. Second, in contrast to construction methods and means for the ontologies of subject areas (which make a special emphasis on the relations among terms and employ artificial intelligence techniques), we consider terminological structure design from the viewpoint of experts in subject areas.

In the context of terminological structure design, two key questions arise immediately. How “well” does a certain set of terms characterize a given subject area and how should one construct the most adequate set of terms? Is it possible to “sort” these terms, e.g., by generality, significance, frequency of occurrence in different word combinations, etc. and how should one do it?

As a matter of fact, this paper almost ignores the first question (except an example in Section 4.1). Concerning the second question, we acknowledge the presence of two extremes in its answer: (1) the expert approach (when an individual or a group of experts in a subject area identify a set of terms and analyze their relations) and (2) the automated approach (when the above-mentioned operations are performed by a computer according to certain algorithms). The benefits and shortcomings of both approaches appear obvious. As ever, the optimal solution is the “golden mean,” i.e., one should combine these approaches by applying the automated expert
approach (when computer-aided methods provide support for expert activity). And we naturally adhere to the combined approach in the present paper.

Suppose that there exists a “thesaurus” (as a rule, created by experts), i.e., a dictionary, a collection of knowledge, a package or code with comprehensive coverage of notions, definitions and terms in a special field of knowledge or a sphere of activity (a subject area). The presence of a thesaurus aims at facilitating correct lexical and corporate communication (simply speaking, good understanding during communication and interaction of individuals belonging to a same discipline or profession); in modern linguistics, a thesaurus is a special type of general- or special-lexis vocabularies with explicitly stated semantic relations among lexical units (synonyms, antonyms, paronyms, hyponyms, hyperonyms, etc.) [32]. Section 2 outlines general approaches to thesauri design and analysis.

In other words, a thesaurus describes a certain subject area by enumerating all its basic notions and their semantic relations. In the elementary form, a thesaurus consists of a list of key terms and semantic (associative and hierarchical) relations among them.

Introduce the following format for a thesaurus of the terminological structure of a subject area (see Fig. 1a). Initially it includes a set of entries for defined terms.

We emphasize that term description possesses wider content than its definition. Moreover, by a generally accepted assumption, the latter is a part of the former, i.e., description totally embraces definition, see Fig. 1a. Therefore, defining terms necessarily appear among describing ones, see Fig. 1c. Defining terms either enter definable terms or lie beyond a studied subject area (e.g., basic philosophical categories).

Fig. 1. A thesaurus of the terminological structure of a subject area: The format and elements.