

Ανάπτυξη της ελληνικής ερευνητικής Υποδομής

για τις Ανθρωπιστικές Επιστήμες ΔΥΑΣ

MIS 441245

ΠΑΡΑΔΟΤΕΟ ΠΑ 3.1 ΙΤΕ 2

Έκθεση αποτελεσμάτων αξιολόγησης καλών πρακτικών.

Σεπτέμβριος 2015

ΙΔΡΥΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΡΕΥΝΑΣ (ΙΤΕ)

ΥΠΟΕΡΓΟ: «ΔΡΑΣΕΙΣ ΑΝΑΠΤΥΞΗΣ ΤΗΣ ΔΥΑΣ»

ΕΠΙΧΕΙΡΗΣΙΑΚΑ ΠΡΟΓΡΑΜΜΑΤΑ : «**ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ & ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ**» ΚΑΙ
«**ΠΕΡΙΦΕΡΕΙΩΝ ΣΕ ΜΕΤΑΒΑΣΗ**»

ΕΘΝΙΚΟ ΣΤΡΑΤΗΓΙΚΟ ΠΛΑΙΣΙΟ ΑΝΑΦΟΡΑΣ

ΕΣΠΑ 2007-2013



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Παιδείας & Θρησκευμάτων
Γενική Γραμματεία Έρευνας & Τεχνολογίας



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ ΑΝΑΠΤΥΞΗΣ & ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑΣ



η περιφέρεια στο **επίκεντρο** της ανάπτυξης

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης - Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης (ΕΤΠΑ), στο πλαίσιο του Ε.Π. Ανταγωνιστικότητα και Επιχειρηματικότητα (ΕΠΑΝ ΙΙ) και των Π.Ε.Π. Αττικής, Π.Ε.Π. Μακεδονίας - Θράκης

ΠΑΡΑΔΟΤΕΟ ΠΑ 3.1 ΙΤΕ 2

Έγγραφο:	DARIAH ΚΡΗΤΗ ΠΑ 3.1 ΙΤΕ 2
Τίτλος παραδοτέου	ΈΚΘΕΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΑΞΙΟΛΟΓΗΣΗΣ ΚΑΛΩΝ ΠΡΑΚΤΙΚΩΝ
Ενότητα Εργασίας:	Συντονισμός με δραστηριότητες της ευρωπαϊκής υποδομής DARIAH
Υπεύθυνος Φορέας/μονάδα:	ΙΤΕ – ΙΠ
Μονάδα	Εργαστήριο Πληροφοριακών Συστημάτων Κέντρο Πολιτισμικής Πληροφορικής
Άλλοι συμμετέχοντες:	
Συγγραφείς:	Μάρτιν Ντέρ, Χρυσούλα Μπεκιάρη, Μαρία Δασκαλάκη, Λήδα Χαραμή
Κατάσταση Εγγράφου:	Α' έκδοση
Ημερομηνία πρώτης έκδοσης	Σεπτέμβριος 2015
Ημερομηνία τελευταίας επικαιροποίησης	
Σχετικοί σύνδεσμοι	

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ 4

ΠΑΡΑΡΤΗΜΑ Α 6

 DEVELOPING A MODEL FOR SUSTAINABLE INTEROPERABLE KNOWLEDGE MAINTENANCE 6

ΠΑΡΑΡΤΗΜΑ Β 8

THESAURUS MAINTENANCE 8

 Our point of view 9

 Criteria and Method 10

 Expected benefits 11

 Deliverables: 11

Περίληψη

Στα πλαίσια του παραδοτέου αυτού το ITE σε συνεργασία με την Ακαδημία Αθηνών έγιναν οι παρακάτω ενέργειες:

(α) το ITE σε συνεργασία με την Ακαδημία Αθηνών πρότεινε στη 4^η Γενική συνάντηση των VCC ([4th General VCC meeting](#)), στην Ρώμη που διεξήχθη στις 7-19 Σεπτεμβρίου του 2014 και στα πλαίσια του VCC3 την δημιουργία μιας ομάδας εργασίας για την διασύνδεση δομημένων λεξιλογίων σε πανευρωπαϊκό επίπεδο. Η πρόταση παρουσιάζεται στο παράρτημα Α.

(β) Η πρόταση αυτή έγινε αποδεκτή από το VCC3. Η τελική πρόταση εφαρμογής παρουσιάζεται στο παράρτημα Β.

(γ) Στα πλαίσια της πανευρωπαϊκής ομάδας εργασίας για την διασύνδεση και συντήρηση των δομημένων λεξιλογίων (TMG: Thesaurus Management Group) που συστήθηκε έγιναν οι παρακάτω συναντήσεις:

1. εναρκτήρια συνάντηση στο Βερολίνο στις 18 Δεκεμβρίου 2014, στο Deutsches Archäologisches Institut. Στην συνάντηση αυτή ο Μάρτιν Ντέρ από το ITE παρουσίασε μια μεθοδολογία για την διασύνδεση των λεξιλογίων. Επίσης συντονιστής του TMG ορίστηκε η Ακαδημία Αθηνών.

Το πρόγραμμα και τα αποτελέσματα της συνάντησης αυτής έχουν αναρτηθεί στους ιστοτόπους του DARIAH EU

<https://wiki.de.dariah.eu/display/DARIAH/WG+Thesaurus+Maintenance>, και στον ιστοτόπο του ΔΥΑΣ - DARIAH KRHTH http://83.212.168.219/DariahCrete/ypiresies_Dariah_3.2.

2. Συμμετοχή στη 5^η Γενική συνάντηση των VCC ([5th General VCC meeting](#)), στο Υπουργείο Παιδείας, Επιστημών και Αθλητισμού της Σλοβενίας, στην Λιουμπλιάνα στις 22-24 Απριλίου 2015. Στην συνάντηση αυτή έγινε η παρουσίαση των εργασιών της ομάδας, και αποφασίσθηκε να γίνει μια διήμερη συνάντηση στο Ηράκλειο της Κρήτης. Στην συνάντηση αυτή συμμετείχαν εκ μέρους του ITE η Μαρία Δασκαλάκη, η οποία ήταν η εισηγήτρια της παρουσίασης.

Η παρουσίαση αυτή έχει αναρτηθεί στους ιστοτόπους του DARIAH EU <https://wiki.de.dariah.eu/display/DARIAH/WG+Thesaurus+Maintenance>, και στον ιστοτόπο του ΔΥΑΣ - DARIAH KRHTH http://83.212.168.219/DariahCrete/ypiresies_Dariah_3.2

3. Οργάνωση και υλοποίηση της 3^{ης} συνάντησης εργασίας του TMG. Η συνάντηση αυτή έγινε στις 16-17 Ιουνίου, στο ΙΤΕ στο Ηράκλειο της Κρήτης. Στην συνάντηση συζητήθηκαν τα παραδοτέα της ομάδος εργασίας.

Το πρόγραμμα της συνάντησης και οι παρουσιάσεις έχουν αναρτηθεί στο ιστότοπο του DARIAH EU <https://wiki.de.dariah.eu/display/DARIAH/WG+Thesaurus+Maintenance>, και στον ιστότοπο του ΔΥΑΣ - DARIAH KRHTH

http://83.212.168.219/DariahCrete/ypiresies_Dariah_3.2

Επίσης στους προαναφερθέντες ιστοτόπους αναρτώνται και οι ενημερώσεις των εκδόσεων της ομάδας TMG, δεδομένου ότι η δραστηριότητα αυτή συνεχίζεται μέχρι τον Μάιο του 2016.

Παράρτημα Α

Developing a model for sustainable interoperable knowledge maintenance

The aim is to develop a model and a proposal how existing thesauri and ontologies can be made interoperable and be maintained in a sustainable way. Within this framework we intend to focus on the following subjects:

Methods for the development in a collaborative way of semantically interoperable thesauri by an ontology-driven faceted analysis.

Development of a model of the collaborative research activities of building interoperable thesauri and ontologies as an integral part of a DARIAH research methods ontology

Integrating concepts of goal-oriented generic research functions with notions of scholarly primitives and digital research activities.

Developing a model ensuring the validity of knowledge gained under scope (3).

For that purpose:

a) We will collect information about which thematic areas have been covered in which languages by relevant thesauri and ontologies in Europe.

b) We will investigate and classify the semantic principles by which these relevant existing and used thesauri have been built, in order to assess ways and effort to make them interoperable.

c) We will develop the specific ontology driven faceted analysis of thesaurus development, which allows for integrating existing and new systems of terminology in a common concise and manageable framework regardless the size of the systems under consideration

d) We will propose a model of (electronic) collaboration, in which a few high-level concepts and principles are controlled by a centralized procedure, whereas increasing specialization will allow for more community-specific management of agreement on terms. The guiding principle for this hierarchical management will be the degree of coverage/ recall versus specialization/ precision of the concepts under discussion.

e) In parallel, we will develop a model how documenting (tracing) provenance of knowledge, based on goal oriented research functions at any level of granularity can be integrated into a DARIAH

research methods ontology so that that scholars can follow arguments and the degree of truth of scholarly statements in digital media now and in the future.

Proposers to VCC3

Academy of Athens

ICS-FORTH

Παράρτημα Β

Thesaurus Maintenance

In a world of ever-increasing digitization of knowledge and information, the attempts to classify the existing scientific knowledge by thesauri are growing continuously. This process, however, does not just entail a simple conversion of terms to another form, but a task that requires scientific accuracy and consistency inasmuch as the core of this task is to identify concepts that reflect the commitment of each scientific community to what is regarded real and can be part of a common language among different scientific fields.

This task becomes even more complicated when it comes to the humanities, where it is difficult to distinguish the meaning of a concept from its respective interpretative approaches. However, even in the humanities, it is possible to build a provision of definitions for top-level concepts that can be used throughout the entire spectrum of humanities fields regardless of their specialization, through the comparison of their specific definitions in the framework of every specific field of the humanities. Through this comparison of the concepts, it becomes evident that every concept implies some fundamental properties that are necessary for its definition regardless of the specialization of the field, from which every concept emerges. These properties can help finding more comprehensible definitions and will allow for much more consistent semantic structuring of thesauri – even if developed in parallel.

Therefore the questions that concern us in the process of designing and organizing the thesauri for the humanities are similar to general questions facing computer science in planning systems for conveying knowledge:

- How to achieve the greatest possible accessibility and interoperability of the thesauri.
- How to ensure that users can find, learn, understand and use the terms needed for their research inquiries with a reasonable effort and efficiency.
- How to avoid, as far as possible, the subjective semantic organization of the terms, in order to ensure their validity. On other words, how to achieve the validity of the concepts, which can be guaranteed through intersubjectivity.
- How to ensure that all users, regardless of their background and knowledge, will have access via these terms to the information that may have an impact on their research.
- How to achieve the greatest economy in the process of organizing terms in a common framework and a distributed, collaborative environment.

- How to avoid the logical errors or idiosyncratic decisions, which will lead to inconsistencies and consequently would require backwards-incompatible restructuring of classification systems and their application .

Our point of view

Starting from the belief that these questions did not arise as a result of the attempts to digitize knowledge but are a restatement of the classic scholarly claims for objectivity, interdisciplinarity and transparency that govern science since its birth, we aim to identify the top-level-concepts that should become the common basis for designing and establishing an overarching thesaurus for all humanities.

It is important to point out that the definition of the top-level concepts cannot simply be the outcome of a compromise between the various fields in humanities; nor can that be the principles or methods for deciding how to select and define concepts. In order to adopt principles and methods that can lead us to design sustainable thesauri for the humanities, we should have in mind that definitions of the top-level concepts should express their fundamental properties, which are acceptable regardless of the scientific field in which they are applied. But *how is that possible?*

In this effort we exploit all the advantages offered by categorical semantics, which are based on the definition of general concepts (facets and hierarchies), under which we can subsume more specific terms. One of the major advantages of this kind of classification is the potential of the sustainable and manageable expansion of thesauri into new areas of knowledge in which the general terms continue to be effective and efficient. The reason is that the designing of thesauri based on the definition of concepts that are very high in the hierarchy (facets and hierarchies) is not dependent on the specific scope of their application, while at the same time the classification under these top-level concepts can efficiently express the meaning of the terms in a particular field, and still imply the general meaning at the same time. On the other side, it is not possible to develop a specialized terminology without including general terms pertaining to matter, activity, people, location etc. Consequently, many specialized terminologies, being focussed on the application domain, include badly defined and mutually incompatible definitions of the necessary general terms, which make the whole hierarchies under them unnecessarily incompatible. It is important to point out, that the crucial points in this kind of classification are the criteria and the method applied in order to define the top-level concepts.

Criteria and Method

The demand for objectivity and interdisciplinarity in defining the top-level concepts of the thesaurus leads us to search for methodological principles that are generally valid regardless of the subject that adopts them or the field to which they are applied. For this reason we have turned our attention to principles that are more substance-oriented, since the classifications that are context-oriented are dependent on many factors, which are empirical and thus irrelevant to the properties of the concepts/terms themselves (such factors are the subject who classifies, its background, the existing interpretations of the terms, the influences of the social environment, the context, in which the terms are used etc.). Therefore, unlike classification systems that are based on the affinity of the meaning of the terms which implies a sort of dependency on subjective factors (interpretation of the meaning, background of the subjects, etc.) or systems that are seeking objectivity based on external factors that are irrelevant to the concepts themselves (for example systems that are based on the co-occurrence of term use), we want to demonstrate the inner connection between the terms in a way that these terms are not contradictory to each other. The consistency between the terms of the same hierarchy is based on the principle that the fundamental properties we attribute to a term should be implied by all the terms in that hierarchy. The consistency between the terms of the same hierarchy is therefore a typical criterion that is valid regardless of subjective factors or specifications of the relevant scientific field of its application. For instance, X-Ray systems are used by many disciplines, such as medicine, material assaying, art conservation, archaeology. Classifying them as “medical instruments” or “archaeological instruments” would be incompatible, and would not render anything about their nature. In contrast, they are “instruments for structure analysis of solid things” by substance, rather than by use.

The backbone in our efforts to classify the terms according to typical criteria is the ISA relationship, by means of which we can avoid categorical errors resulting from the subsumption of terms under facets or hierarchies, which have properties different than those of the indexed terms. Every hierarchy should therefore obey this principle of subsuming narrower to broader terms so that the internal consistency of the terms will become clear and explicit. The ISA relationship serves therefore as a logical control to avoid contradictions and achieve the objectivity, interdisciplinarity and transparency, which are the basic scientific requests that should be satisfied in the process of designing a thesaurus. In practice, this requires careful control which implicit properties are indeed more comprehensive than others and not mutually competing, which is a demanding, collaborative intellectual effort.

However, although the demand for objectivity, interdisciplinarity and transparency requires the application of logico-typical criteria, we don't underestimate the importance of the conceptual closeness of the terms. For this reason, in the thesauri declare the affinity of the meaning between the terms by means of "related term"-relationships.

Expected benefits

We can summarise the advantages of our methods in the designing of a thesaurus as follows:

1. It enables the definitions of the top-level-concepts (facets and hierarchies) in a consistent and objective way.
2. It enables the classification, which is valid and transparent regardless of the subjective factors and specialisations of the relevant scientific field to which it is applied.
3. It ensures that users can find, learn, understand and use the terms needed for their research inquiries with reasonable ease and efficiency.
4. It constitutes a methodological guidance for users to create new compatible with a wider or even global framework.
5. It enables collaboration, cross-disciplinary resource discovery, detection of common principles, simplification and enrichment of methods with elements from other disciplines.
6. It can be expanded without the danger of inconsistencies and backwards-incompatible restructuring.
7. It enables the compatibility with other thesauri that are restricted to particular areas of knowledge.
8. It enables the possibility of designing a multilingual and multidisciplinary thesaurus.

Deliverables:

1. Open invitation for a meeting between the teams that design the thesauri under DARIAH, in order to investigate, in more detail, the problems and possibilities of designing multilingual and multidisciplinary thesauri.
2. Organize two Workshops aiming to establish a common understanding between the participants of the way we proceed and the method we follow in order to define the top-level concepts that are fundamental in the process of classifying the terms in thesauri relating to the humanities.

3. Accurate determination of approximately 55 top-level concepts, which are necessary in planning and organizing a multidisciplinary and multilingual thesaurus for the humanities. We will also provide a number of guide terms to be used within each facet.