

## Mondoc: Informal Lightweight Ontology for Faceted Semantic Classification of Hypernymy

**Authors :** M. Regina Carreira-Lopez

**Abstract :** Lightweight ontologies seek to concrete union relationships between a parent node, and a secondary node, also called "child node". This logic relation (L) can be formally defined as a triple ontological relation (LO) equivalent to LO in (LN, LE, LC), and where LN represents a finite set of nodes (N); LE is a set of entities (E), each of which represents a relationship between nodes to form a rooted tree of (LN, LE); and LC is a finite set of concepts (C), encoded in a formal language (FL). Mondoc enables more refined searches on semantic and classified facets for retrieving specialized knowledge about Atlantic migrations, from the Declaration of Independence of the United States of America (1776) and to the end of the Spanish Civil War (1939). The model looks forward to increasing documentary relevance by applying an inverse frequency of co-occurrent hypernymy phenomena for a concrete dataset of textual corpora, with RMySQL package. Mondoc profiles archival utilities implementing SQL programming code, and allows data export to XML schemas, for achieving semantic and faceted analysis of speech by analyzing keywords in context (KWIC). The methodology applies random and unrestricted sampling techniques with RMySQL to verify the resonance phenomena of inverse documentary relevance between the number of co-occurrences of the same term (t) in more than two documents of a set of texts (D). Secondly, the research also evidences co-associations between (t) and their corresponding synonyms and antonyms (synsets) are also inverse. The results from grouping facets or polysemic words with synsets in more than two textual corpora within their syntagmatic context (nouns, verbs, adjectives, etc.) state how to proceed with semantic indexing of hypernymy phenomena for subject-heading lists and for authority lists for documentary and archival purposes. Mondoc contributes to the development of web directories and seems to achieve a proper and more selective search of e-documents (classification ontology). It can also foster on-line catalogs production for semantic authorities, or concepts, through XML schemas, because its applications could be used for implementing data models, by a prior adaptation of the based-ontology to structured meta-languages, such as OWL, RDF (descriptive ontology). Mondoc serves to the classification of concepts and applies a semantic indexing approach of facets. It enables information retrieval, as well as quantitative and qualitative data interpretation. The model reproduces a triple tuple (LN, LE, LT, LCF L, BKF) where LN is a set of entities that connect with other nodes to concrete a rooted tree in (LN, LE). LT specifies a set of terms, and LCF acts as a finite set of concepts, encoded in a formal language, L. Mondoc only resolves partial problems of linguistic ambiguity (in case of synonymy and antonymy), but neither the pragmatic dimension of natural language nor the cognitive perspective is addressed. To achieve this goal, forthcoming programming developments should target at oriented meta-languages with structured documents in XML.

**Keywords :** hypernymy, information retrieval, lightweight ontology, resonance

**Conference Title :** ICCLCL 2020 : International Conference on Cognition, Language and Computational Linguistics

**Conference Location :** Rome, Italy

**Conference Dates :** July 23-24, 2020