TERMINOLOGY, TRANSLATION AND ARTIFICIAL INTELLIGENCE

Clinton P. Smyth - Advisor, Minerva Intelligence Inc., Vancouver, Canada
Tatjana Krzman - Manager of Translation Services, MOSAIC, Vancouver, Canada

ABSTRACT

Many of the nouns or noun phrases humans use can be arranged into taxonomies where one noun/noun phrase denotes something that is a kind of something denoted by another noun/noun phrase. For example, a cow is a kind of animal. Knowledge of these taxonomic relationships is critical to correct reasoning, whether it be carried out by a human or by a computer. Unfortunately, taxonomic relationships between noun/noun phrases are often either confused with meronomic (part-of) relationships, or the two are conflated into one hierarchical tree structure, as allowed by the widely-adopted Simple Knowledge Organisation System (SKOS).

To address this problem, workers in artificial intelligence have developed applications to assist in distinguishing between taxonomies and meronomies, and in expressing taxonomies as graphs rather than as trees in order to implement multiple-inheritance of properties during reasoning. The successful deployment of these applications relies on careful analysis of the definition of each noun/noun phrase in a taxonomy. Consequently, in the context of language translation, these applications can also play an important role in exposing different meanings carried by nouns or noun phrases in different languages which are expected to have the same meaning. Such differences, if not resolved, will result in faulty transfer of knowledge in, and provision of explanations by, artificial intelligence systems that operate in more than one language.

1 Introduction

Early in the book "An Introduction to Language" (Fromkin, Rodman, and Hyams 2018) , the authors include the following quotation from Noam Chomsky's seminal study "Language and the Mind" (Chomsky 2006):

> *"When we study human language, we are approaching what some might call the "human essence," the distinctive qualities of mind that are, so far as we know, unique to man."*

This paper addresses some of the challenges in expressing some of that human essence in machines where it can be used to discover and explain new knowledge.

As will become apparent in the paper, it is helpful to define what we mean by "knowledge" in this context. For the purposes of this paper, our definition of knowledge will be "something that is known and can be written down". This kind of knowledge is called "explicit knowledge" to distinguish it from "tacit knowledge", learned only by experience, and communicated only indirectly, through metaphor and analogy (Nonaka and Takeuchi 1995).

Indeed, another way of expressing the purpose of this paper is to say that it addresses some of the challenges of making implicit knowledge explicit so that computers can simulate human reasoning.

2 Words and their Purpose

Words may be considered the smallest unit of meaning used to communicate knowledge or information between human beings and computers. Phrases, sentences, paragraphs, chapters and books may be considered larger information communication units.

Words carry meaning for those who send them out by writing or speaking, and they are intended to carry meaning to those receiving them. Depending on circumstances, the meaning understood by the receiving party may be exactly the same as that understood by the sending party, or close to, or significantly different from the sender's understanding. For a detailed exposition on words and what they are understood to mean in the context of artificial intelligence (AI), see Section 13.2 entitled "Symbols and Semantics" in the online book "Artificial Intelligence: Foundations of Computational Agents" (Poole and Mackworth 2017).

Much of human beings' common understanding of words is derived from their shared experience of life. Many "understandings", however, of rarer or more complex concepts, are learnt by humans from questioning each other, or by reference to dictionaries and related resources where the meanings of words are defined or described.

We may make four important observations at this stage:

1. Different life experiences or reference sources may lead different individuals to understand different things from the same word (mining engineer's understanding of "dump site" may be different from that of a town planner);
2. Different words have evolved in different languages to denote the same concept ("forest" in English is "šuma" in Bosnian, or "Wald" in German) which in many instances indicates evolution from the proto language;

3. When we carefully consider the definitions or descriptions of words from different languages which are considered to represent the same concept, we may find significant difference in the words' intended meaning;
4. Computers are not humans, so cannot have life experience, and can therefore use words only in ways they have been programmed to do by human beings.

An entertaining example of (3) above is provided by (Deutscher 2011):

> Culture: civilisation, the state of being cultivated, refinement, the result of cultivation, a type of civilisation (Chambers English Dictionary)
>
> Kultur: Gesamtheit der geistigen und künstlerischen Errungenschaften einer Gesellschaft (*The totality of intellectual and artistic achievements of a society*.) Störig German Dictionary

Related to observation (3) above is the fact that certain terms in one language may have no equivalent in a second language, as situation commonly referred to as a "lexical gap". Lexical gaps are discussed in detail in (Hann 2004).

Importantly for expressing knowledge in words on computers, researchers have shown that it is frequently words denoting higher-level, more general concepts (see 5.1 and 6.1 below), that do not correspond directly across language, and therefore are more difficult to translate, with translators having to use adjectives, and sometimes long phrases, to accurately create a translation. A quick and easy solution that some governments and even linguistic authorities sometimes resort to is adopting words from the language(s) where a certain concept exists and introducing them to the general public through media or literature. Examples of such words can be found in the Serbian language (e.g. "management" = "менаџмент" or "tender" = "тендер"). This process of adoption, unfortunately, is not always accompanied by proper adaptation, which brings us back to the observation (3 above).

3 Translation and its Purpose

Translation is an act through which the content of a text is transferred from the source language in to the target language (Foster 1958). The purpose of the translation is usually to convey factual information, the nature of which can be inferred from the definitions of the words used in the translation. Alternatively, it may be primarily to convey a feeling, or a sentiment, which is not mentioned in the expression being translated. Two translations below of the German poem by Heinrich Heine serve as an example. The first pays more attention to conveying the information content of the original and the second to conveying emotion (as discussed in detail in (Deutscher 2011)).

Ein Fichtenbaum steht einsam Im Norden auf kahler Höh'. Ihn schläfert; mit weißer Decke Umhüllen ihn Eis und Schnee. Er träumt von einer Palme, Die, fern im Morgenland, Einsam und schweigend trauert Auf bren nender Felsenwand. Heinrich Heine	A pine-tree standeth lonely In the North on an upland bare; It standeth whitely shrouded With snow, and sleepeth there. It dreameth of a Palm Tree Which far in the East alone, In mournful silence standeth On its ridge of burning stone. (Translator: James Thompson)	There stands a lonely pine-tree In the north, on a barren height; He sleeps while the ice and snow Swathe him in folds of white flakes. He dreameth of a palm-tree Far in the sunrise-land, Lonely and silent longing On her burning bank of sand. (Translator: Emma Lazarus)

While translation has historically been between human beings, during the last many decades it has also come to apply to communication between humans and computers, and equally importantly in the present day, to communication between computers independently of humans. Computer-to-computer translation may be required in the same human language, as, for example, is required for concurrent querying of computer databases and maps of the same type but which use different classification systems (eg: Land Use Classification). This requirement is typically called "interoperability" in the computer sciences, and is typically met by implementation of a "mediator". A mediator typically provides mapping between any number of different classification systems and an overarching classification system designed to cater for as many classes as practicable chosen from the different systems. Practicability in this context is determined by the system designers and their perception of future user needs. Just as such mediator systems need the right domain expertise to map between classification systems which are all in the same language, with the right cross-language expertise, they can be designed to mediate between classification systems deployed in different languages.

One example of such a system delivering an international coastal atlas is described in (Lassoued et al. 2008) and shown graphically in Figure 1.

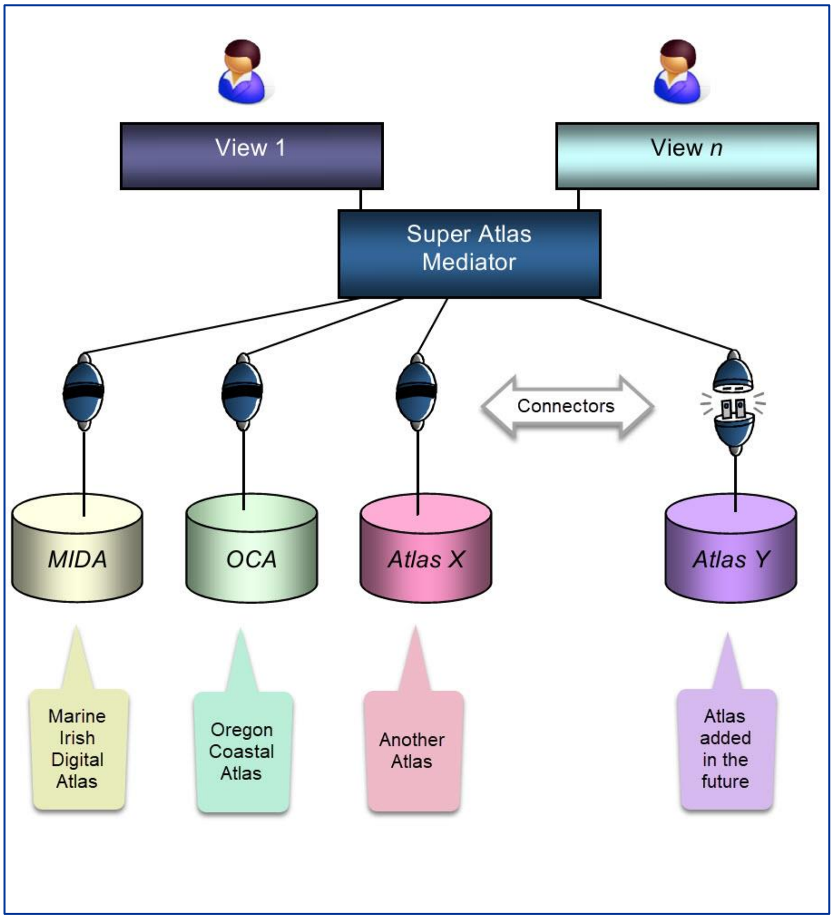


Figure 1: The place of mediators in providing interoperability between maps which use different terminologies. In a virtual integration architecture such as mediation, atlases can be added or removed quite easily without affecting the super atlas, provided that they have the right connectors (Adapted from (Lassoued et al. 2008)).

Mediators typically require that the terminologies they relate to each other are each described in their own ontology. Ontologies, discussed in section 6.1 below are knowledge structures which make explicit the meaning of each word in the terminology they represent, a pre-requisite for accurate cross-referencing of different terminologies, whatever their language. Figure 2 below shows the relationships between ontologies that typically support a mediator.

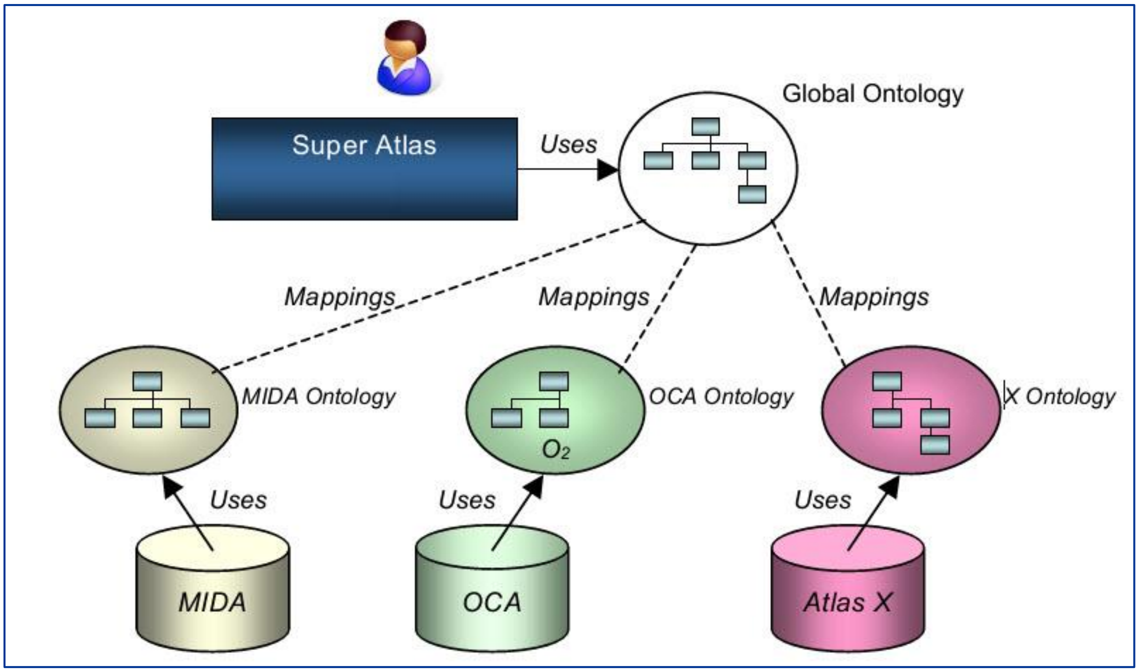


Figure 2: The various ontologies required for development of a cross-terminologies mediator (from (Lassoued et al. 2008)).

The disciplined multi-language terminology mappings required for interoperability of cognitive AI system require a lot of work by persons who are experts in the domain of the AI system (for example, law or geology) as well as translators with a deep knowledge of the source and target languages. This work is greatly facilitated by fit-for-purpose software tools. Certain relevant software, such as ontology and taxonomy editors for OWL have been designed to operate in a very broad range of languages. Other important tools, such as "ShowVoc" and "Synonym Finder" are currently under development in the European Union, and are currently available for use in early versions (Publications Office of the European Union 2022) and (European Commission 2022).

Figure 3 overleaf illustrates ShowVoc's rendition of search results for the term "anthropology" as carried out over all the terms in the ELSST Thesaurus. There are four "useability" points to note in Figure 3:

I. No definition of the word "anthropology" is provided;
II. Although the language filter was set to display only English, Spanish and Croatian, the filter has been applied only to the "Label" fields, resulting in a crowded screen showing many languages not relevant to the user's query;
III. Though Croatian translations were requested, none are available, which is true for a number of the other languages catered for by ShowVoc. Clearly, much translation work remains to be done in this important context.
IV. Preferred and Alternative Labels are shown twice – probably a simple programming error.

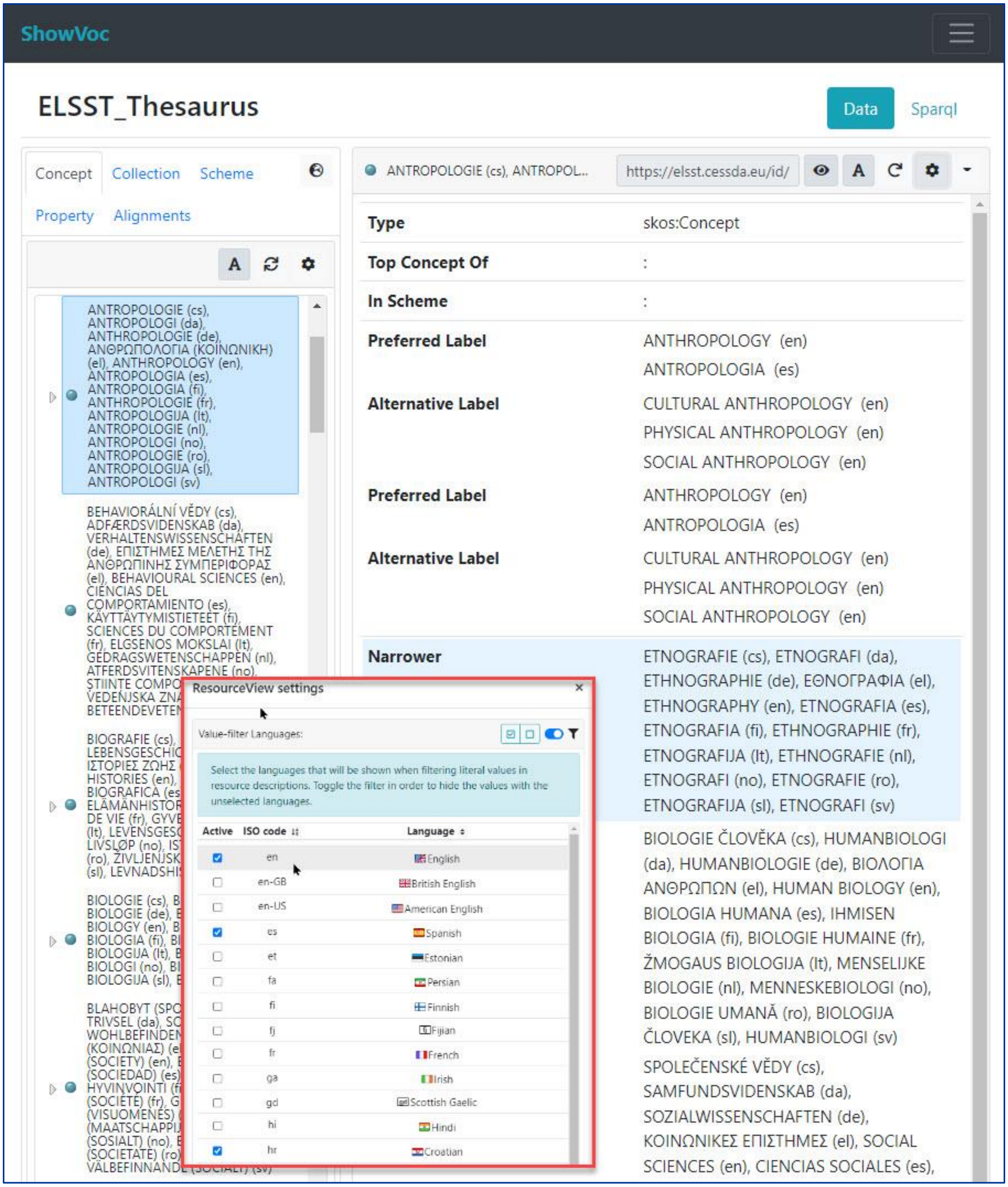


Figure 3: Results from searching the ELSST Thesaurus for the term "anthropology" through the ShowVoc web service with the language filter set to show only English, Spanish and Croatian languages (Publications Office of the European Union 2022).

The purpose of Alternative Labels in ShowVoc (see Figure 3) is to record synonyms, knowledge of which humans use all the time when talking and reading. For computer programs to use them (as in text searches and cognitive AI applications) they need to be stored somewhere in a format the computers can be programmed to use. Because of their importance to data and document discoverability, synonyms have become a subject of interest to the European Commission which , together with KU Leuven, recently published the report entitled “Using Synonyms to better Data Discoverability” (European Commission. Joint Research Centre. and KU Leuven. 2022). At the same time they released a web application called “Synonyms Finder” which automatically collects and can display synonyms from a number of internet-accessible resources such as FAO, INSPIRE and WikiData vocabularies (European Commission 2022). In Figure 4 Synonyms Finder displays the synonyms it harvested from EuroVoc, WikiData and Eionet resources for the concept “railway line”. To test the language interoperability of Synonyms Finder Figure 4 was generated with the language switched from English to Spanish in the language selection drop-down on the search page.

Significantly, while this choice of Spanish changed some of the text on the page to Spanish, it did not provide any links to Spanish translations of the synonyms displayed – although programmatically it is not difficult to do so for those resources which store their vocabularies in more than one language, such as WikiData (WikiData 2022). Figure 5 shows the WikiData page for the concept “railway line” which includes its definition as well as its translation into the three languages of interest to the WikiData user (as configured on their “Preferences” page). Also on the displayed page, of significant relevance to developers of multi-lingual cognitive AI applications, is a definition of each translated term, and a declaration, according to WikiData, of three sub-classes of the concept “railway line” and one “part-of” a railway line. Because WikiData covers data in a great many languages, a side-bar on the page provides links to additional information on the concept “railway line” in an additional eleven languages.

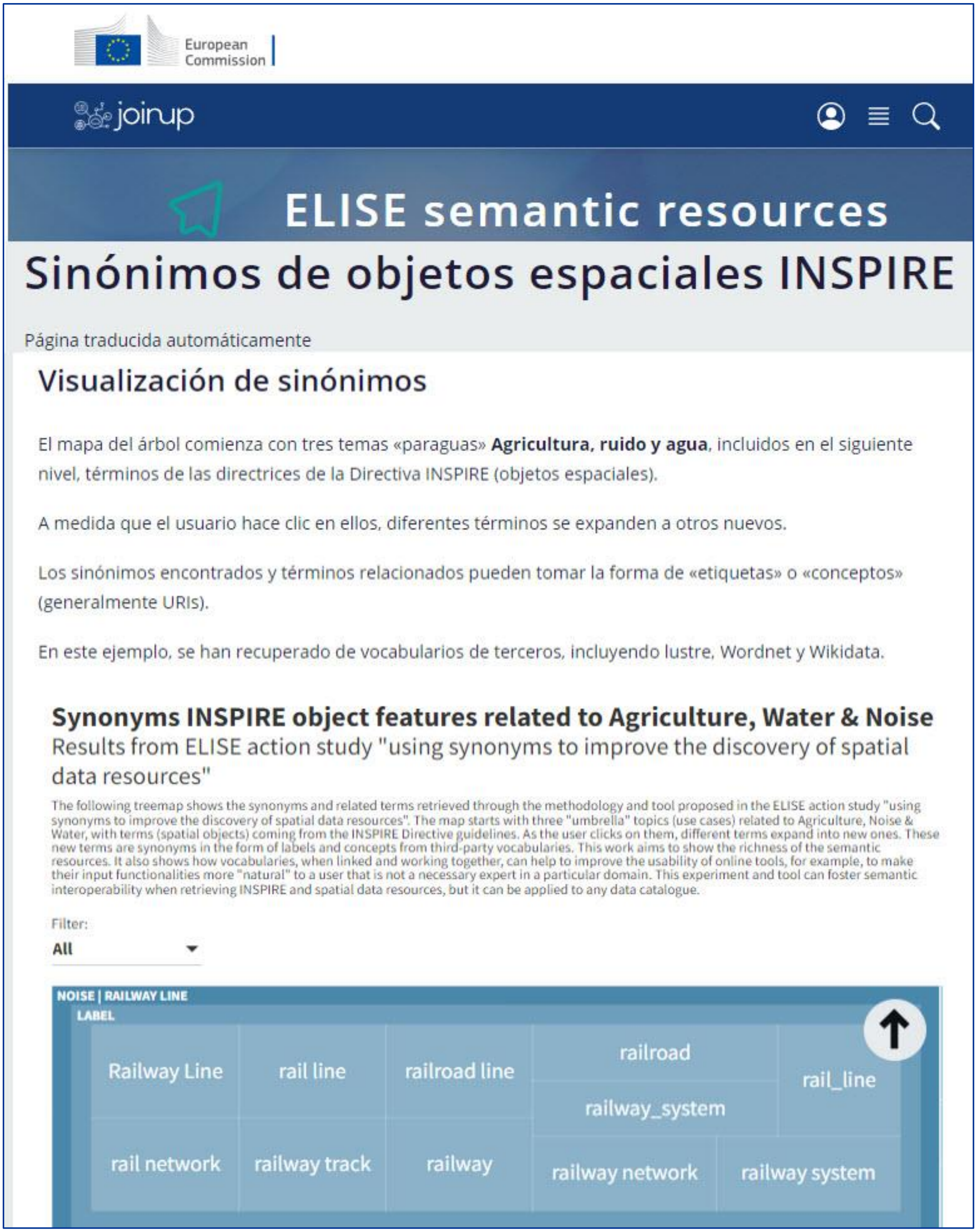
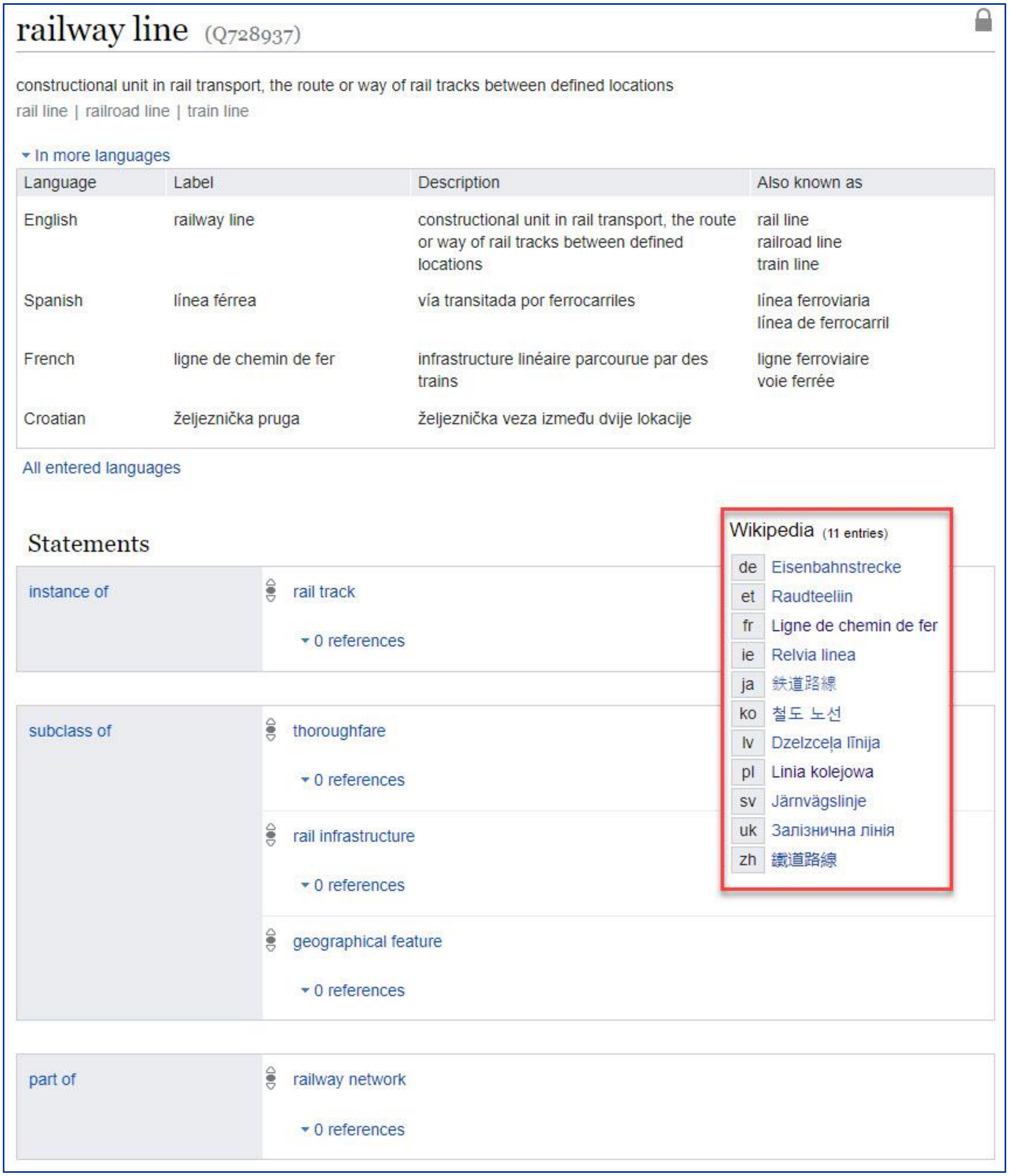


Figure 4: Condensed screenshot of the EU's "Synonyms Finder" application showing synonyms for "railway line" harvested from EuroVoc, WikiData and Eionet, with language requested set to Spanish (European Commission 2022).



railway line (Q728937)

constructional unit in rail transport, the route or way of rail tracks between defined locations
rail line | railroad line | train line

In more languages

Language	Label	Description	Also known as
English	railway line	constructional unit in rail transport, the route or way of rail tracks between defined locations	rail line railroad line train line
Spanish	línea férrea	vía transitada por ferrocarriles	línea ferroviaria línea de ferrocarril
French	ligne de chemin de fer	infrastructure linéaire parcourue par des trains	ligne ferroviaire voie ferrée
Croatian	željeznička pruga	željeznička veza između dvije lokacije	

All entered languages

Statements

instance of
rail track
0 references

subclass of
thoroughfare
0 references
rail infrastructure
0 references
geographical feature
0 references

part of
railway network
0 references

Wikipedia (11 entries)

de	Eisenbahnstrecke
et	Raudteeliin
fr	Ligne de chemin de fer
ie	Relvia linea
ja	鉄道路線
ko	철도 노선
lv	Dzelzceļa līnija
pl	Linia kolejowa
sv	Järnvägslinje
uk	Залізнична лінія
zh	鐵道路線

Figure 5: WikiData page showing information about the concept "railway line" (WikiData 2022).

4 Artificial Intelligence

Artificial intelligence, or AI, is the field that studies the synthesis and analysis of computational agents that act intelligently. This definition is from the book "Artificial Intelligence: Foundations of Computational Agents" (Poole and Mackworth 2017), from which much of the following section is paraphrased.

An agent is something that acts in an environment; it does something. Agents include worms, dogs, thermostats, computers, robots, humans, companies, and countries. AI is the study of intelligent behaviour in computational terms.

4.1 Kinds of AI

There are many aspects to, and categories of, AI.

Two major categories are those of Machine Learning (ML) and Cognitive AI. Machine Learning involves algorithms (possibly with bootstrapping training data) that learn from (typically large sets of) data to create information or carry out tasks. Examples are speech recognition, Google Translate and weather forecasts.

Cognitive AI strives to simulate intelligent human thought and behaviour. It typically involves analysis of the real-world environment, context, intent and many other variables that inform a person's ability to solve problems. The fundamental differences between ML and Cognitive AI are discussed in detail in the book "Rebooting AI: Building Artificial Intelligence We Can Trust" (Marcus and Davis 2019).

Cognitive AI and its need for the translation of language expressing human knowledge into language useable in by computers programmed to simulate intelligent human behaviour lies at the heart of this paper. This new language needed by cognitive computers is understandably emerging to be very similar to existing natural languages, but is much more strongly standardised, and expected to be that way for a long time.

While there are many ways to test for intelligent behaviour in cognitive AI systems, Levesque (Levesque 2014) has posed question-answering as fundamental in this regard, with an example following his Winograd schema being the following:

- The city councilmen refused the demonstrators a permit because they feared violence. Who feared violence?
- The city councilmen refused the demonstrators a permit because they advocated violence. Who advocated violence?

These two sentences differ only in one word feared/advocated, but have the opposite answer. Answering such a question depends on knowing something about the language describing the world that humans understand, but computers currently do not.

Winograd schemas have the property that (a) humans can easily disambiguate them and (b) there is no simple grammatical or statistical test that could disambiguate them.

Levesque reports the following conclusions:

1. Much of what we come to know about the world and the people around us is not from personal experience, but is due to our use of language.

   People talk to us, we listen to weather reports and to the dialogue in movies, and we read: text messages, sport scores, mystery novels, etc.

   And yet, it appears that we need to use extensive knowledge to make good sense of all this language.

2. Even the most basic child-level knowledge seems to call upon a wide range of logical constructs.

   Cause and effect and non-effect, counterfactuals, generalized quantifiers, uncertainty, other agents' beliefs, desires and intentions, etc.

   And yet, symbolic reasoning over these constructs seems to be much too demanding computationally.

4.2 Explainable AI

Discussion accompanying the increasingly common deployment of intelligent systems in application domains such as autonomous vehicles and transportation, medical diagnosis, or insurance and financial services have shown that when decisions are taken or suggested by automated systems, it is essential for practical, social, and—with increasing frequency—legal reasons that an explanation can be provided to users, developers, and regulators (Confalonieri et al. 2021).

Confalonieri et al go on to identify seven key considerations to be taken into account for the development of Explainable AI (XAI) systems, namely: Causality, Counterfactuals, Social Context, Selectivity, Transparency, Semantics and Interactivity. In regard to semantics, they declare "If explanations are symbolically grounded—by means of ontologies, conceptual networks, or knowledge graphs—they can support common-sense reasoning. Formal representation and reasoning can in turn enact various forms of knowledge manipulation, such as abstraction and refinement [to simplify explanations, as appropriate]".

An example of explanation output by a cognitive AI system is presented in Section 7.3.

5 Concepts and Categorisation

Concepts are the building blocks of thoughts (Margolis and Laurence 2021). Consequently, they are crucial to such psychological processes as categorization, inference, memory, learning, and decision-making.

In information science in general, and AI in particular, concepts are described using ontologies. In this context, an ontology is a specification of the meaning of the symbols used in an information system, where symbols refer to things that exist (Poole and Mackworth 2017). Ontologies are discussed in greater detail in Section 6.1.

An upper ontology is an ontology which consists of very general terms (such as "object", "property", "relation", "class", "sub-class") that are common across all knowledge domains. An important function of an upper ontology is to support broad semantic interoperability among a large number of domain-specific ontologies by providing a common starting point for the formulation of definitions ("Upper Ontology" 2022).

Categorizing objects, the basis for modern ontologies, has a long history (Aristotle BC350). Aristotle suggested the definition of a class C in terms of:

I. Genus: a superclass of C
II. Differentia: the attributes that make members of the class C different from other members of the superclass of C (and therefore members of a sub-class of C).

The correct categorisation of things is critical to reasoning with the words that represent them. Of particular importance in this regard is the distinction between kind-of and part-of relationships between concepts, as discussed below.

5.1 Types of Classification

5.1.1 FOLK TAXONOMIES AND FOLKSONOMIES

Wikipedia describes folk taxonomies as follows ("Folk Taxonomy" 2021):

> *"A folk taxonomy is a vernacular naming system, as distinct from scientific taxonomy. Folk biological classification is the way people traditionally describe and organize their natural surroundings/the world around them, typically making generous use of form taxa like "shrubs", "bugs", "ducks", "fish" and the like, or of economic criteria such as "game animal" or "pack animal".*
>
> *Folk taxonomies are generated from social knowledge and are used in everyday speech. They are distinguished from scientific taxonomies that claim to be disembedded from social relations and thus more objective and universal. Folk taxonomies exist to allow popular identification of classes of objects, and apply to all areas of human activity. All parts of the world have their own systems of naming local plants and animals. These naming systems are a vital aid to survival and include information such as the fruiting patterns of trees and the habits of large mammals."*

The distinction described above between folk taxonomies and scientific taxonomies is comprehensively examined in the book "Naming Nature – The Clash between Instinct and Science" (Yoon 2009). As evidence of the scope and richness of the folk taxonomies the book "Landmarks" documents hundreds of vernacular landscape-related terms from the United Kingdom, together with their definitions (Macfarlane 2016). Figure 6 provides twelve examples of "Woods and Woodlands-related" terms. Most of these are unlikely to find their way into the kind of global ontology shown in Figure 2. In regard to scientific taxonomies, Wüster et al provide a recent review of "taxonomic vandalism" in biology and the challenges facing scientific taxonomy in general (Wüster et al. 2021). Reconciling advances in knowledge which generates a need for new terms (See Section 8.2 below) with the requirement for stability in taxonomy and nomenclature is a balance their community has to manage.

Word	Location	Meaning
faschboil	Irish	underwood; grove or bosket
frith	Sussex	holy wood; young underwood growing beside hedges
ghost	forestry	destroyed wood whose outline remains as a hedge, soil-mark or boundary
grout	Suffolk	small grove
bagg	Yorkshire	copse or woodland, especially on a slope or hillside
bagginblock	Northern Ireland	wooded area
bake	Ireland	to steal apples
banger	Berkshire, Hampshire	wood on the side of a steep hill or bank
bolt	Cotswalds	high wood
burst	forestry	isolated wood, especially one on a hill
leaf-whelmed	poetic	in such dense foliage that sight is extremely limited
leah	Old English	permanent glade or clearing in woodland

Figure 6: Twelve terms related to various folk taxonomies, past and present, in the United Kingdom (Macfarlane 2016).

Folksonomies, on the other hand, are described as “classification systems [which have arisen on the internet] in which end users apply public tags to online items, typically to make those items easier for themselves or others to find later” (“Folksonomy” 2021).

The term “foldksonomies” is sometimes used derisively in reference to poorly-structured technical or scientific classification schemes which purport to be taxonomies but which mix “kind-of” (taxonomic) and “part-of” (meronomic) relationships into the same data structures without discriminating the two. SKOS (see Section below) makes it particularly easy to combine taxonomic and meronomic relationships into data structures which cannot be used by cognitive AI applications.

5.1.2 TAXONOMY

Taxonomies are structures connecting types via subtyping, i.e., type specialization relations. These structures are fundamental for conceptual domain modeling, and have a central organizing role in areas such as knowledge representation, ontology engineering and computer reasoning.

A characteristic critical to both human and computer reasoning involving a logically-correct taxonomy is that all properties of a parent taxonomic class are inherited by all of its sub-classes. In other words, properties propagate “down” taxonomies, as illustrated for selected levels of the taxonomy of rocks in Figure 10.

Human brains can draw inferences on a great many taxonomic relations almost instantaneously during discussion, reading, thought and reasoning. For example, if one is told that a man was hit by a Ford Mustang, one knows that a person, with all the attributes of a human being, was hit by a car, with all the attributes of a vehicle. Computers can carry out this reasoning only if they have access to a correct taxonomy of vehicles and a correct taxonomy of humans.

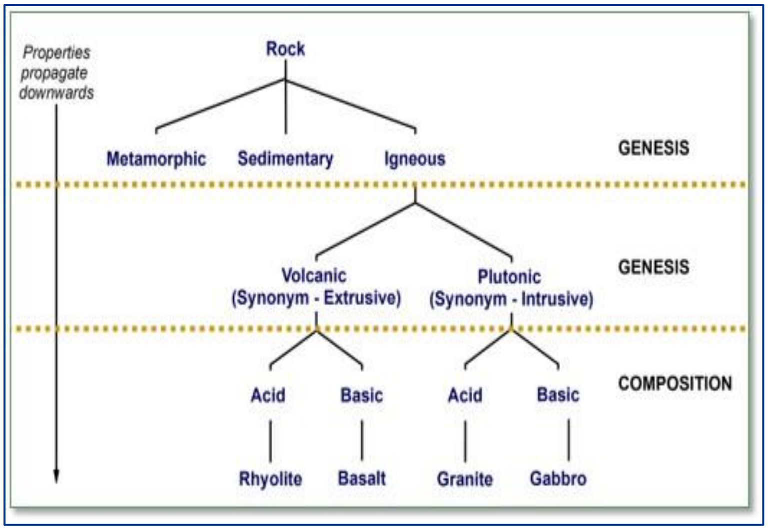


Figure 7: Selected levels in the taxonomy of rocks illustrating that properties propagate down the hierarchy.

Such taxonomies have in recent years assumed great importance in document-search applications (Semantic Web Company 2021) to the extent that the purveyors of ontology and document management companies are partnering with companies which maintain both public and proprietary taxonomies for their clients (Semantic Web Company 2022).

As computer interoperability has become more important in the 21st century, so too has the development of open terminology standards for use in computer data and knowledge bases. The European Union, through its INSPIRE legislation, has been a leader in this field, having published open access terminologies (called "code lists") covering 34 different themes ranging from human health to geology (European Commission-JRC 2007).

Significantly, many of the published terminologies are intended to be taxonomies facilitating powerful querying of maps and other information artefacts. Equally significantly, all the published code lists are available in all the 23 languages of the EU. Figure 8 shows the HILUCS Land Use code list displayed in both English and Croatian (European Commission-JRC 2022).

Unfortunately for translators the European Commission does not yet make their INSPIRE code lists available in translation tables, leaving that work to be undertaken by, and sometimes shared by, other parties.

Figure 9 shows an extract of OpenStreetMap's translation of CORINE Land Cover terms into Bosnian, together with English-based OpenStreetMap terms corresponding to CORINE terms (OpenStreetMap 2022). INSPIRE has adopted the CORINE Land Cover taxonomy as the legislated land cover code list for the European Union (Copernicus 2022). Very detailed descriptions of the intended meaning of each CORINE class has been published in human-readable form by the W3C organisation from an underlying SKOS-formatted rendition of the standard (W3C 2015).

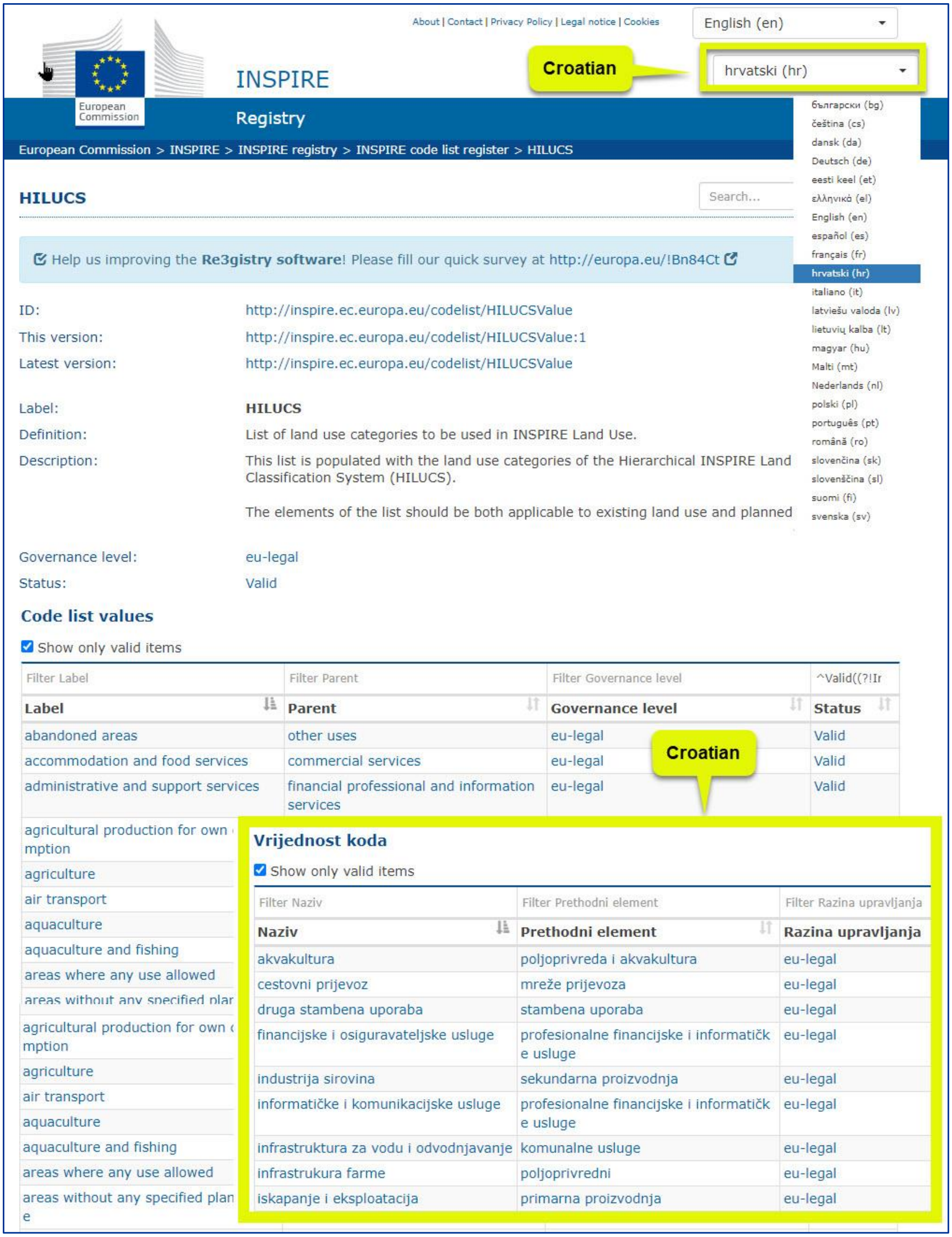
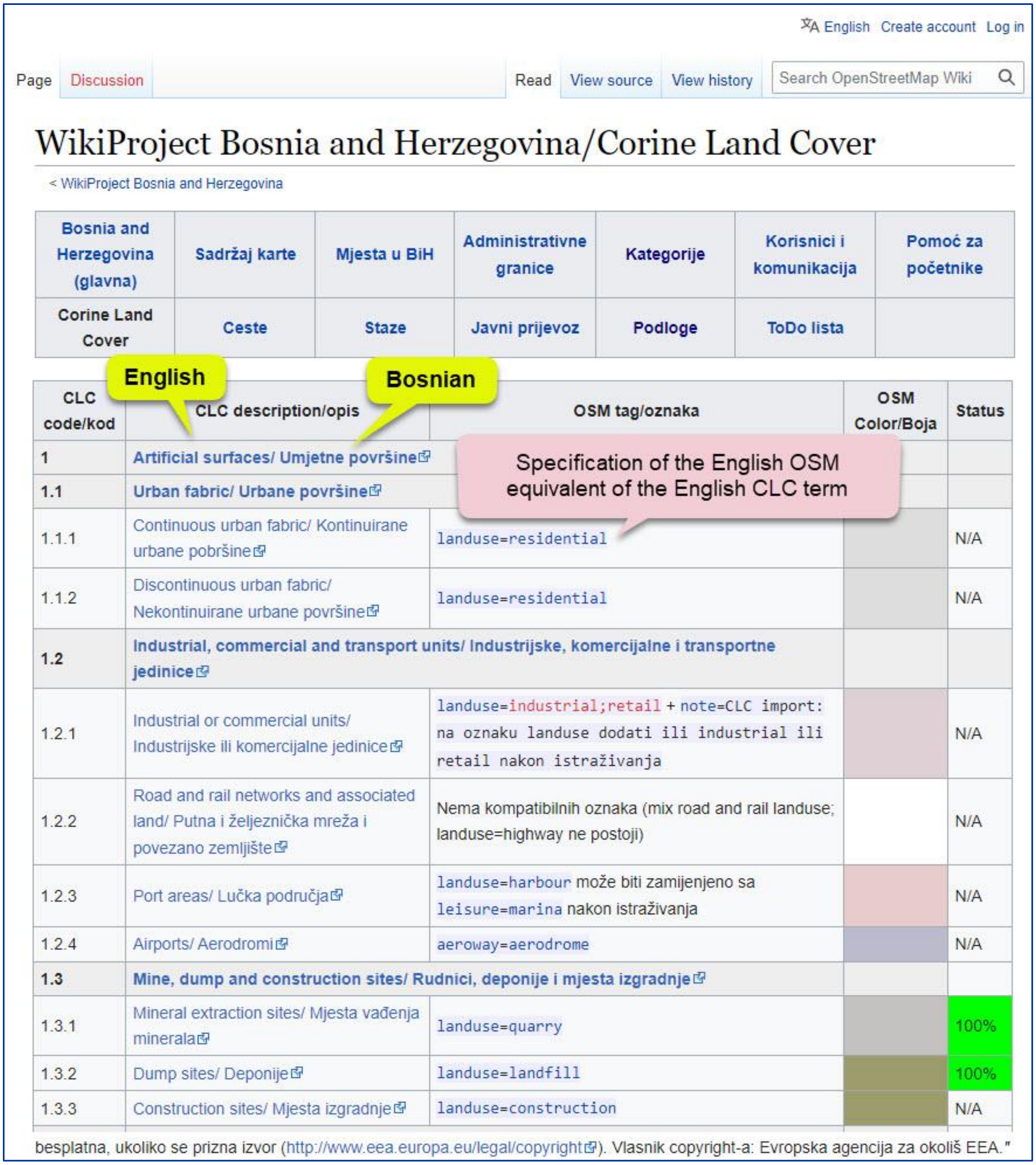


Figure 8: The INSPIRE Land Use (HILUCS) code list as available for download in both English and Croatian on the INSPIRE web site.



文A English Create account Log in

Page Discussion Read View source View history Search OpenStreetMap Wiki

WikiProject Bosnia and Herzegovina/Corine Land Cover

< WikiProject Bosnia and Herzegovina

Bosnia and Herzegovina (glavna)	Sadržaj karte	Mjesta u BiH	Administrativne granice	Kategorije	Korisnici i komunikacija	Pomoć za početnike
Corine Land Cover	Ceste	Staze	Javni prijevoz	Podloge	ToDo lista	

CLC code/kod	CLC description/opis	OSM tag/oznaka	OSM Color/Boja	Status
1	**Artificial surfaces/ Umjetne površine**			
1.1	**Urban fabric/ Urbane površine**			
1.1.1	Continuous urban fabric/ Kontinuirane urbane pobršine	`landuse=residential`		N/A
1.1.2	Discontinuous urban fabric/ Nekontinuirane urbane površine	`landuse=residential`		N/A
1.2	**Industrial, commercial and transport units/ Industrijske, komercijalne i transportne jedinice**			
1.2.1	Industrial or commercial units/ Industrijske ili komercijalne jedinice	`landuse=industrial;retail` + `note=CLC import: na oznaku landuse dodati ili industrial ili retail nakon istraživanja`		N/A
1.2.2	Road and rail networks and associated land/ Putna i željeznička mreža i povezano zemljište	Nema kompatibilnih oznaka (mix road and rail landuse; landuse=highway ne postoji)		N/A
1.2.3	Port areas/ Lučka područja	`landuse=harbour` može biti zamijenjeno sa `leisure=marina` nakon istraživanja		N/A
1.2.4	Airports/ Aerodromi	`aeroway=aerodrome`		N/A
1.3	**Mine, dump and construction sites/ Rudnici, deponije i mjesta izgradnje**			
1.3.1	Mineral extraction sites/ Mjesta vađenja minerala	`landuse=quarry`		100%
1.3.2	Dump sites/ Deponije	`landuse=landfill`		100%
1.3.3	Construction sites/ Mjesta izgradnje	`landuse=construction`		N/A

besplatna, ukoliko se prizna izvor (http://www.eea.europa.eu/legal/copyright). Vlasnik copyright-a: Evropska agencija za okoliš EEA."

Figure 9: Extract of OpenStreetMap translation of CORINE Land Cover terms into Bosnian, together with English-based OpenStreetMap terms corresponding to CORINE terms (OpenStreetMap 2022).

5.1.3 PARTONOMY/MERONOMY

A partonomy or meronomy is a type of hierarchical information structure that expresses part–whole relationships – which are very different from class-subclass relationships.

There is a rich literature on partonomies, particularly in the medical field, where, for cognitive AI systems reasoning about anatomy, the distinction between parts and kinds is critical (Johansson and Lynøe 2013). An arm is a kind of a limb but an elbow is a part of an arm. Brown provides a rich analysis of anatomical partonomy terms and how the have developed from folk taxonomies (Brown 1976).

The earth sciences provide another context for the important partonomic relationship between rocks, minerals and elements, as shown in Figure 10.

An important characteristic of partonomies is that some properties may propagate "up" the partonomic hierarchy. In Figure 10 below, if Mineral 2 contains arsenic, then the rock of which it is a part also contains arsenic.

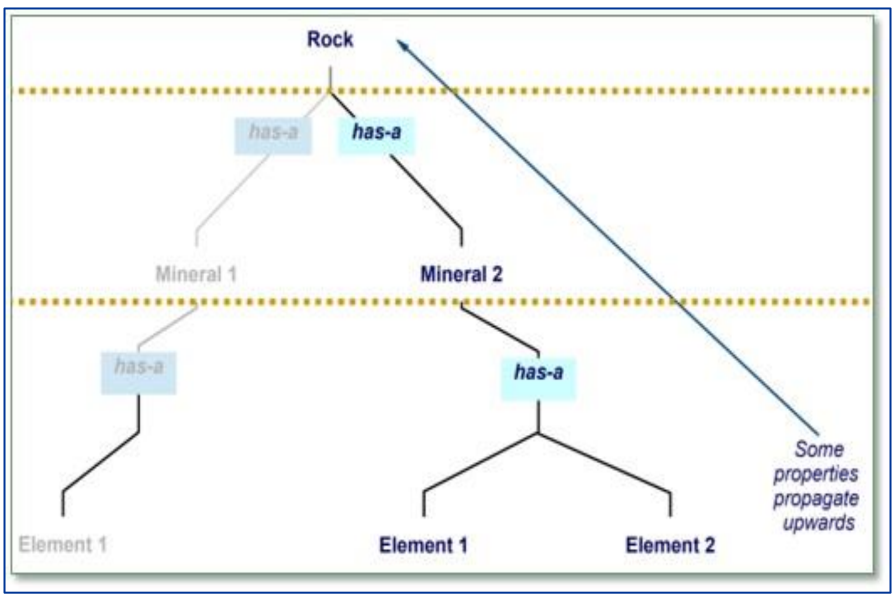


Figure 10: The partonomic relationship between rocks, minerals and elements.

5.1.4 SKOS

SKOS ("Simple Knowledge Organization System") is a W3C recommendation designed for representation of thesauri, classification schemes, subject-heading systems, or various other types of structured controlled vocabulary.

While SKOS can be used to store and represent taxonomies, since its primary hierarchical relationship is defined as the reflexive "is narrower than / is broader than" relationship, it does not conveniently distinguish between "kind-of" and "part-of" relationships, both of which are specialisations of "is narrower than". This has the unfortunate consequence that although a collection of discipline-specific terms may have been arranged into a SKOS-compliant structure, it cannot be used for most reasoning tasks in a cognitive AI system because of the different propagation behaviour of properties in taxonomies and partonomies, as discussed in Sections 5.1.2 and 5.1.3.

6 Knowledge Representation

6.1 Ontologies

Having defined an ontology as a specification of the meaning of the symbols used in an information system in Section 5 above, we here describe typical ontologies used in AI and other systems in more detail.

A detailed discussion of Upper Level Ontologies is beyond the scope of this paper, albeit that they provide the foundational concepts upon which ontology editors are built. For a concise history of the development of upper level ontologies see the paper "Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story" (Guizzardi et al. 2015). Section 8.1.1 below reports a method for developing ontologically correct taxonomies built upon the Unified Foundational Ontology.

The most widely-used ontology language for commercial and research applications is the Web Ontology Language, commonly known as "OWL" (W3C 2012). A typical ontology for a domain of interest (such as farming) will, as made possible by a language like OWL, describe the domain in terms of:

- **Individuals**, being the things in the world that are being described (such as farms and animals);
- **Classes**, which are sets of individuals. A class is the set of all real or potential things that would be in that class. For example, the class "Cow" may be the set of all animals that would be classified as a cow, not just those cows that exist in the domain of interest;
- **Properties**, which are used to describe individuals or entities. For example, properties in the ontology of cows may be that they are all mammals and all have an age.

Like most modern ontology languages, OWL is based on description logic (Poole and Mackworth 2017). This makes it possible to deduce or induce new knowledge from an ontology combined with various individual or entity descriptions made using the terminology defined in that ontology (Hogan et al. 2022). In other words, ontologies make possible reasoning with existing explicit information or knowledge to produce, or make explicit, knowledge which was before hidden and implicit. In so doing ontologies, and their building blocks, such as taxonomies and partonomies, are fundamental to engineering cognitive AI applications.

6.2 Knowledge Graphs and Semantic Networks

As usefully defined by Hogan et al, "a knowledge graph is a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between entities" (Hogan et al. 2022).

Aided by the development of the internet and graph database technology, knowledge graphs have evolved from semantic networks (Wikipedia 2022) and concept graphs (Sowa 2005) which were intended to express small quanta of information (equivalent to sentences or paragraphs) to graphs with many millions of nodes and edges, generally without any equivalent of punctuation in their unprocessed state. There are many ways to extract knowledge from large graph databases.

Knowledge graphs have emerged as a compelling abstraction for organizing the world's structured knowledge and for integrating information extracted from multiple data sources (Chaudhri et al. 2022). They are also beginning to play a central role in representing information extracted by AI systems, and for improving the predictions of AI systems by giving them ontologically-controlled knowledge expressed in knowledge graphs as input.

Figure 11 presents a simple semantic network describing an imagined mineral deposit (Sharma, Poole, and Smyth 2010).

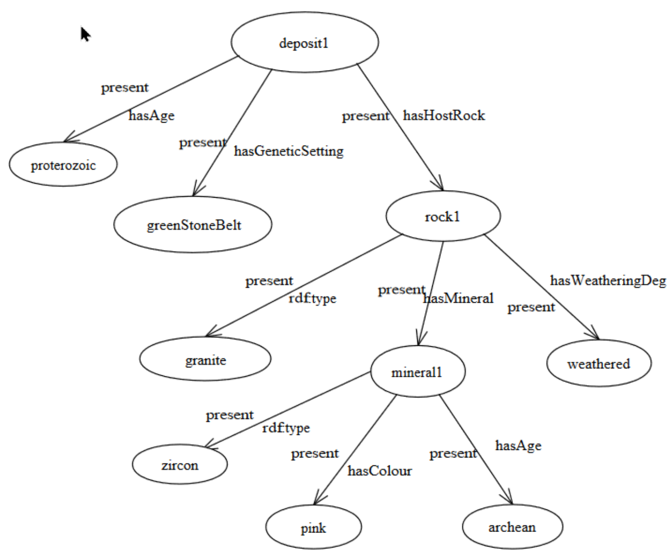


Figure 11: A simple semantic network describing an imagined mineral deposit.

By contrast, Figure 12 presents a knowledge graph showing the relationships between the characters in "Harry Potter and the Philosopher's Stone" (Bratanic 2021). Wikidata's knowledge graph is orders of magnitude greater than this knowledge graph.

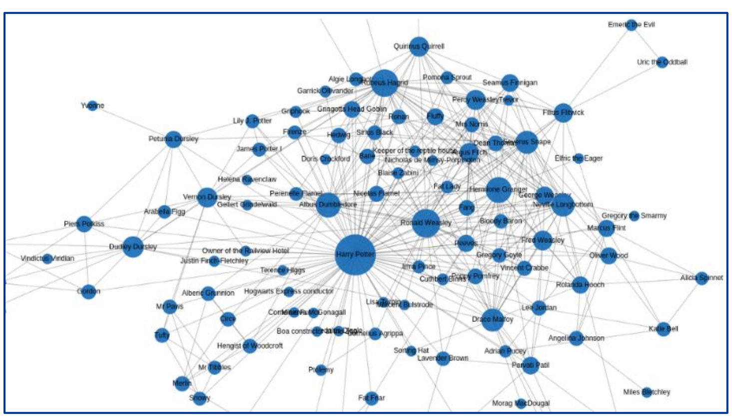


Figure 12: the relationships between the characters in "Harry Potter and the Philosopher's Stone" (Bratanic 2021).

7 Cognitive AI, Similarity Rankings and Explainable AI

Cognitive science links various disciplines that study cognition and reasoning, from psychology to linguistics to anthropology to neuroscience. AI distinguishes itself within cognitive science by providing tools to build intelligence rather than just studying the external behavior of intelligent agents or dissecting the inner workings of intelligent systems (Poole and Mackworth 2017).

7.1 Similarity

One of the hallmarks of cognition is the ability to recognise similarity between concepts and entities, and to rank them on similarity when there are many comparisons to be made. In the words of Goldstone and Son (Goldstone and Son 2005):

> *"Human assessments of similarity are fundamental to cognition because similarities in the world are revealing. The world is an orderly enough place that similar objects and events tend to behave similarly."*

In the following section we briefly present two fielded cognitive AI systems that are built around the principles of comparison and similarity ranking.

7.2 Similarity Ranking

7.2.1 LANDSLIDE HAZARD MAPPING AND MINERALS EXPLORATION

Roberti et al describes a landslide susceptibility mapping system, built using the semantic standards of INSPIRE, which ranks the level of landslide susceptibility in different areas of the Italian province of Veneto by comparing them with semantic descriptions of various types of landslide (Roberti et al. 2020).

In the system, each different landslide type (or model) is described with a simple semantic network and compared with the thousands of different discrete areas (map polygons) in Veneto, each described with a semantic network to produce the kind of susceptibility map shown in Figure 13. Figure 14 shows one simplified model being compared with one simplified area description.

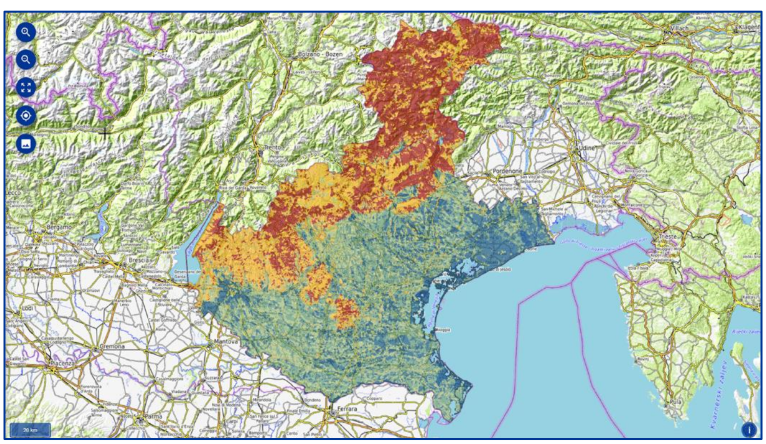


Figure 13: Web map interface showing susceptibility to slides in soil in Veneto, Italy (Roberti et al. 2020).

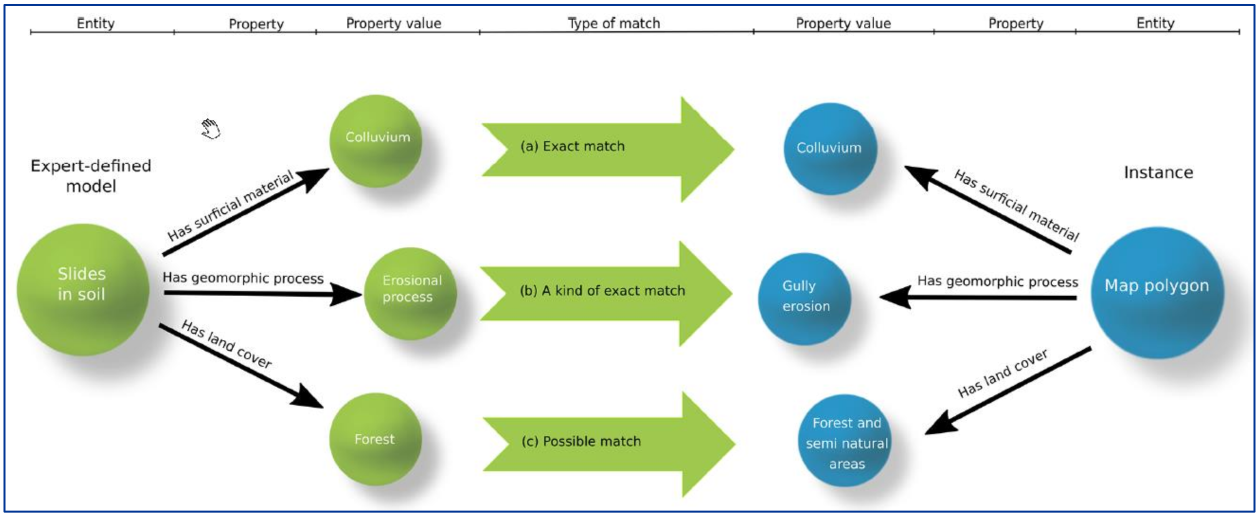


Figure 14: Graphical representation of the matching process between expert-defined models and map polygon instances. See text for explanation.

In Figure 14, "Type of match" (a) is an example of an exact match between the property value "colluvium" in both semantic networks; (b) is an example of a kind of (AKO) exact match because "gully erosion" is a more specific kind of "erosional process". The model is looking for an "erosional process" and found a "gully erosion"; (c) is an example of a possible exact match because "forest and semi-natural areas" is a broader concept of "forest". The model is looking for "forest", but we do not know whether the instance is a "forest". We only know that the instance is "forest or a semi-natural area". The vocabulary's taxonomic relationships are supplied to the reasoning code in the AI application by the ontology.

Detailed consideration of the weights applied to matching or conflicting properties is beyond the scope of this paper, and is explained in (Roberti et al. 2020). Of note, however, is the fact, obvious to any human being capable of cognitive reasoning, that relatively higher weightings should be, and are, given to "kind of exact" matches ((b) above) and lower weightings should be given to possible matches (such as (c) above).

A similar approach to that described by Roberti et al for landslides susceptibility mapping has been adopted by Minerva Intelligence Inc. for minerals exploration targeting in the Yukon. The resulting system, which was also built primarily on INSPIRE terminology standards is available for review at https://minervaintelligence.com/target/ . It is noteworthy that the INSPIRE standards used in this system were originally developed by the IUGS Commission for the Management and Application of Geoscience Information (https://cgi-iugs.org/) under the name of "GeoSciML".

7.2.2 CORPORATE ENVIRONMENT, SOCIAL AND GOVERNANCE (ESG) REPORTING

In an article dated 16 April, 2022, one of Canada's leading national newspapers reported that "The world of finance is facing a reckoning over what defines sustainable investing" and they produced the information shown in Figure 15 below to illustrate how very differently six different companies were rated by seven different rating agencies (Jones and Milstead 2022). These are important ratings as they influence where large investing institutions invest their money.

In anticipation of the need for such rankings, and the requirement that standard vocabularies be used by companies reporting on their ESG performance every quarter (if meaningful comparisons are to be made), the European Community published its "EU Taxonomy for Sustainable Activities" in 2020. They describe the taxonomy as follows:

> *"The EU taxonomy is a classification system, establishing a list of environmentally sustainable economic activities. It could play an important role helping the EU scale up sustainable investment and implement the European green deal. The EU taxonomy would provide companies, investors and policymakers with appropriate definitions for which economic activities can be considered environmentally sustainable. In this way, it should create security for investors, protect private investors from greenwashing, help companies to become more climate-friendly, mitigate market fragmentation and help shift investments where they are most needed."*

Same company, different ESG rating

The Globe's investigation of ESG ratings showed a wide divergence among some Canadian companies. The Globe compared ESG ratings of companies in the S&P/TSX composite index from six different providers and roughly assigned them into higher (green), middle (yellow) and lower (red) tiers.

Higher Middle Lower

Company	Bloomberg	S&P	MSCI	Sustainalytics	CDP Climate	ISS	ISS
Ritchie Bros. Auctioneers	28.5	18	AAA	N/A	N/A	4	20.28
FirstService Corp.	13.6	15	A	N/A	0	8	17.45
Toromont Industries	24.0	9	AA	N/A	0	6	27.66
AltaGas	52.9	23	A	N/A	7	5	27.94
BCE	59.7	54	BBB	18.9	7	1	59.50
IGM Financial	53.1	74	BBB	11.7	7	8	49.25

How the ratings work

Bloomberg ESG disclosure
Scores range from 0 (no information provided) to 100 (all possible information provided)

The S&P Global ESG rank
A percentile score with 1 being the worst and 100 being the best

The MSCI ESG rating
The system ranges from best (AAA) to worst (CCC)

Sustainalytics rank
0-10 Negligible; 10-20 Low; 20-30 Medium; 30-40 High; 40+ Severe

CDP Climate score
Scores range from 0 (failure) to 8 (A)

ISS Governance scores
Range from 1 (best) to 10 (worst)

ISS ESG Performance scores
Range from 1 (worst) to 100 (best)

THE GLOBE AND MAIL, SOURCE: BLOOMBERG, ISS, GLOBE RESEARCH

Figure 15: Table showing a wide divergence of ESG ratings of six Canadian companies as rated by seven ratings agencies (Jones and Milstead 2022).

This is clearly another domain in which standard terminology, available in all EU languages at least, will play an important role in computer-assisted societal discourse. Society should expect assistance from computers in all three of the following related respects: (1) from running the scoring algorithms ratings agencies are already using to generate their ESG rankings from parsed quarterly reports (2) from making those scoring systems transparent and public and (3) by making each score generated explainable in natural language using the technologies described in this paper.

7.3 Explainable AI Revisited

AI has entered the business mainstream, opening up opportunities to boost productivity, innovation and fundamentally transform operating models (PWC 2018). Explainable AI has emerged as a response to the "black box" problem of AI, according to which models and their performance are not understandable by humans. The need for explanations being built in to AI applications has been comprehensively reviewed by Gerlings et al (Gerlings, Shollo, and Constantiou 2021).

It is clear that explainability requires that AI developers work with subject-matter experts to engineer software that can generate human-comprehensible explanations. It is also regularly pointed out that there is a generally inverse relationship between the performance of an AI method and its explainability, as shown in Figure 16. Performance in this context can broadly be thought of as "degree of sophistication". The inescapable fact remains, however, that explanations, be they of any level of sophisticated AI engineering, will have to be made in language with well-defined terms exactly matching the concepts they reference, no matter what the language they are delivered in.

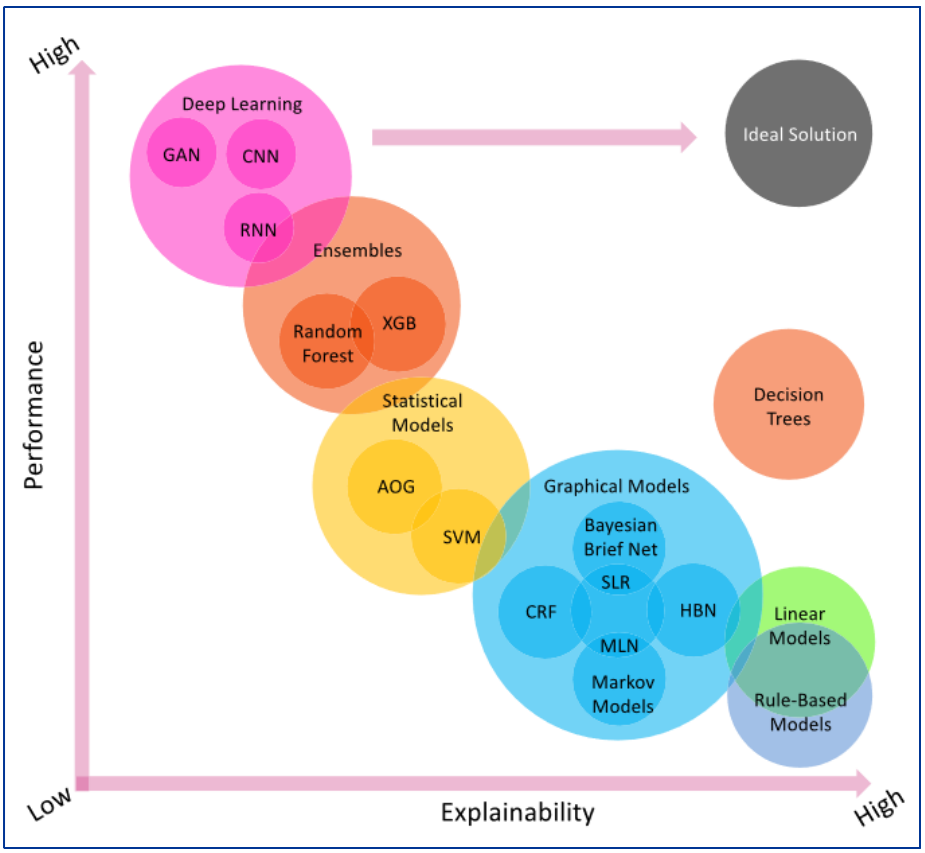


Figure 16: Model explainability vs. model performance for widely used machine learning and deep learning algorithms. The ideal solution should have both high explainability and high performance. However, existing linear models, rule-based models and decision trees are more transparent, but with lower performance in general (Yang, Ye, and Xia 2022).

Even the simple example of AI explainability provided by IBM in a 2021 video (IBM 2021) demonstrating explainability in AI makes clear the need for clarity in the meaning of terms. A summary of the explanation of the credit risk evaluation result discussed in the IBM video is presented in Figure 17 below.

If readers of the explanation do not understand what the bank means by “checking status” or “others on loan”, or do not have access to their definitions, they will not understand the explanation.

Credit Risk Model X					
No Risk			**CONFIDENCE**		**Risk**
	44%			**55%**	
Factors contributing to No Risk				**Factors contributing to Risk**	
Attribute	**Value**			**Attribute**	**Value**
Others on Loan	None	10%			
			20%	Age	51
Loan duration	15	9%			
			17%	Checking status	none
Telephone	none	4%			
			15%	Employment duration	>7
Sex	male	3%			
			9%	Current residence duration	4
			6%	Owns property	car
			4%	Credit history	Outstanding credit

Explanation:	**Credit Risk Model X predicts RISK with 55% confidence.**	
	The following features were the most important in determining this prediction:	
	Age	**51**
	Checking status	**none**
	Employment duration	**>7**

Figure 17: Summary of an explanation generated from an AI-derived credit risk rating published by IBM (IBM 2021), making clear the terms, such as "checking status", used in the explanation that need to be understood by the explainee.

8 Ontologically Correct Taxonomies

8.1 The Role of Logic

Because of the critical role taxonomies play in reasoning (see Section 5) much work has been done on software tools for creating and maintaining logically correct taxonomies.

Two notable examples of this work are the taxonomy editor called ACE (Minerva Intelligence 2020) and the recently available seminal paper entitled “Ontologically Correct Taxonomies by Construction” (Batista et al. 2022).

8.1.1 ACE: THE ARISTOTELIAN CLASS EDITOR

The purpose of the ACE taxonomy editor, which is available for free use on the internet, is to provide a tool for converting a draft vocabulary (for example, a folksonomy or a folk taxonomy) consisting of terms and text definitions into a logically coherent ontology for use in AI computation. The intended primary user is a domain expert with a clear understanding of the Aristotelian approach to class definition. The main ACE workflow is as follows:

1. Analyzing definitions to identify the differentiating properties of each class as attribute/value pairs;
2. Establishing the required values each attribute may assume which may themselves be hierarchical;
3. Assigning all distinguishing property values to each term in the target taxonomy;
4. Using the OWL Reasoner built in to ACE to infer from their properties all the class/sub-class relationships in the taxonomy;
5. Evaluate the resulting taxonomy, which typically has a graph rather than a tree topology, for consistency with expert opinion, adjusting sub-class properties and re-running the taxonomy inference if errors are identified.

The final result is a complete, conceptually-valid taxonomy which can be used for reasoning by a cognitive AI program.

In a 2022 paper, Batista et al have further formalised the development of logically correct taxonomies by invoking upper level ontology concepts and implementing the OntoClean methodology for taxonomy generation (Batista et al. 2022). The authors report that a plug-in for the popular free-to-use Protégé ontology editor is being developed which will generate “Ontologically Correct Taxonomies by Construction” using their methodology (Stanford University 2022).

8.2 The Role of Translators in developing taxonomies for Explainable AI

Both the above methods for development of logically correct taxonomies depend on accurate identification of the properties (attribute/value pairs) that define and distinguish classes and sub-classes.

Great care will need to be taken in the development of Explainable AI applications which need to provide their explanations in more than one language. This is because, as we have seen in Sections 2 to 5 above, many concepts exist which are represented in different languages by words which have very similar, but not exactly the same, meaning. Yet explanations output by AI programs need to be exact in their meaning, at least in some transparent auditable way (as by the easy provision of word or phrase definitions).

A solution to this kind of problem sometimes lies in the use of adjectives to qualify a noun in one language to align it with a noun in another language. However, one can imagine situations where a new word would need to be used for all languages in order not to create confusion. Whatever the case, it appears certain that the drive to impart computers with intelligence has highlighted the need to be very explicit about what words mean when they are input into computers.

It is going to be the work of translators to ensure that when domain experts are “engineering” knowledge into computers, they are doing it in a way that will allow the computers’ explanations to be understandable in any of the languages of their future users. A good place to start is by developing taxonomies that use words that make sense withing the taxonomy in all relevant languages.

9 Conclusion

Cognitive AI needs the translation of language expressing human knowledge into language useable by computers programmed to simulate intelligent human thinking, reasoning and behaviour.

This new language needed by cognitive computers is understandably emerging to be very similar to existing natural language, but is much more strongly standardised, and expected to be that way for a long time.

It is going to be the work of translators to ensure that when domain experts are “engineering” knowledge into computers, they are doing it in a way that will allow the computers’ explanations to be understandable in any of the languages of their future users.

A good place to start in this endeavour is by developing taxonomies, using available public-domain software tools, that use words that make sense within the taxonomy in all relevant languages.

10 References

Aristotle. BC350. *Aristotle’s Categories and de Interpretatione*. Edited by J. L. Ackrill. Oxford: Oxford University Press.

Batista, Jeferson O., João Paulo A. Almeida, Eduardo Zambon, and Giancarlo Guizzardi. 2022. “Ontologically Correct Taxonomies by Construction.” *Data & Knowledge Engineering* 139 (May): 102012. https://doi.org/10.1016/j.datak.2022.102012.

Brown, Cecil H. 1976. “General Principles of Human Anatomical Partonomy and Speculations on the Growth of Partonomic Nomenclature.” *American Ethnologist* 3 (3): 400–424. https://doi.org/10.1525/ae.1976.3.3.02a00020.

Chaudhri, Vinay K, Chaitanya Baru, Naren Chittar, Xin Luna Dong, Michael Genesereth, James Hendler, Aditya Kalyanpur, et al. 2022. “Knowledge Graphs: Introduction, History, and Perspectives.” *AI MAGAZINE* 43: 17–29. https://doi.org/10.1002/aaai.12033.

Chomsky, Noam. 2006. *Language and Mind*. 3rd edition. Cambridge ; New York: Cambridge University Press.

Confalonieri, Roberto, Ludovik Coba, Benedikt Wagner, and Tarek R. Besold. 2021. “A Historical Perspective of Explainable Artificial Intelligence.” *WIREs Data Mining and Knowledge Discovery* 11 (1). https://doi.org/10.1002/widm.1391.

Copernicus. 2022. “CORINE Land Cover — Copernicus Land Monitoring Service.” Land Section. 2022. https://land.copernicus.eu/pan-european/corine-land-cover.

Deutscher, Guy. 2011. *Through the Language Glass: Why the World Looks Different in Other Languages*. Illustrated edition. New York: Picador.

European Commission. 2022. “Synonyms Finder.” 2022. https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/solution/elise-semantic-resources/synonyms-finder.

European Commission. Joint Research Centre. and KU Leuven. 2022. *Using Synonyms to Better Data Discoverability: Application to INSPIRE Spatial Objects.* LU: Publications Office. https://data.europa.eu/doi/10.2760/08796.

European Commission-JRC. 2007. “INSPIRE.” 2007. https://inspire.ec.europa.eu/.

———. 2022. “HILUCS.” 2022. https://inspire.ec.europa.eu, inspire.ec.europa.eu/codelist/HILUCSValue.

“Folk Taxonomy.” 2021. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Folk_taxonomy&oldid=1041979009.

“Folksonomy.” 2021. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=1046799344.

Fromkin, Victoria, Robert Rodman, and Nina Hyams. 2018. *An Introduction to Language*. 11th edition. Boston, MA: Wadsworth Publishing.

Gerlings, Julie, Arisa Shollo, and Ioanna Constantiou. 2021. “Reviewing the Need for Explainable Artificial Intelligence (XAI).” In . https://doi.org/10.24251/HICSS.2021.156.

Goldstone, Robert L., and Ji Yun Son. 2005. “Similarity.” In *The Cambridge Handbook of Thinking and Reasoning*. Cambridge University Press. http://library.mibckerala.org/lms_frame/eBook/Holyoak-Morrison%20-%20The%20Cambridge%20Handbook%20of%20Thinking%20and%20Reasoning%20(CUP).pdf.

Guizzardi, Giancarlo, Gerd Wagner, João Paulo Andrade Almeida, and Renata S.S. Guizzardi. 2015. “Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story.” *Applied Ontology* 10 (3–4): 259–71. https://doi.org/10.3233/AO-150157.

Hann, Michael. 2004. *A Basis for Scientific and Engineering Translation: German-English-German*. Bilingual edition. Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia D’amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, et al. 2022. “Knowledge Graphs.” *ACM Computing Surveys* 54 (4): 1–37. https://doi.org/10.1145/3447772.

IBM. 2021. *Explainable AI Video*. https://www.ibm.com/watson/explainable-ai.

Johansson, Ingvar, and Niels Lynøe. 2013. *Medicine & Philosophy: A Twenty-First Century Introduction*. Walter de Gruyter.

Jones, Jeffrey, and David Milstead. 2022. “Why the Booming Business of ESG Ratings May Be Giving Investors a False Sense of Sustainability.” *The Globe and Mail*, April 16, 2022. https://www.theglobeandmail.com/business/article-behind-greenwashing-esg-ratings-sustainable-investing/.

Lassoued, Y., D. Wright, L. Bermudez, T. Nyerges, Tanya Haddad, and N. Dwyer. 2008. “Semantic Mediation as a Gateway to Interoperability, with a Case Study of the International Coastal Atlas Network (ICAN).” In *Undefined*. Park City, UT, USA: Springer-Verlag. https://www.semanticscholar.org/paper/Semantic-Mediation-as-a-Gateway-to-with-a-Case-of-Lassoued-Wright/a67e469108b1c487e3e1ad5660775a11a6d5ada8.

Levesque, Hector J. 2014. “On Our Best Behaviour.” *Artificial Intelligence* 212 (July): 27–35. https://doi.org/10.1016/j.artint.2014.03.007.

Macfarlane, Robert. 2016. *Landmarks*. Illustrated edition. London: Penguin Books.

Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. Vintage.

Margolis, Eric, and Stephen Laurence. 2021. “Concepts.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/concepts/.

Minerva Intelligence. 2020. “ACE - The Aristotelian Calss Editor.” Minerva Intelligence. 2020. https://minervaintelligence.com/ace/.

Nonaka, Ikujiro, and Hirotaka Takeuchi. 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Illustrated edition. New York: Oxford University Press.

OpenStreetMap. 2022. “WikiProject Bosnia and Herzegovina/CORINE.” 2022. https://wiki.openstreetmap.org/wiki/WikiProject_Bosnia_and_Herzegovina/Corine_Land_Cover.

Poole, David L., and Alan K. Mackworth. 2017. *Artificial Intelligence: Foundations of Computational Agents*. 2 edition. Cambridge, UK: Cambridge University Press. https://artint.info/2e/html/ArtInt2e.html.

Publications Office of the European Union. 2022. “ShowVoc.” 2022. https://showvoc.op.europa.eu/#/home.

PWC. 2018. “Explainable AI: Driving Business Value through Greater Understanding.” Pricewaterhouse Coopers. https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf.

Roberti, Gioachino, Jacob McGregor, Sharon Lam, David Bigelow, Blake Boyko, Chris Ahern, Victoria Wang, et al. 2020. “INSPIRE Standards as a Framework for Artificial Intelligence Applications: A Landslide Example.” *Natural Hazards and Earth System Sciences* 20 (12): 3455–83. https://doi.org/10.5194/nhess-20-3455-2020.

Semantic Web Company. 2021. “What Is Taxonomy Management?” https://www.poolparty.biz/wp-content/uploads/2021/04/White-Paper-What-is-Taxonomy-Management.pdf.

———. 2022. “Semantic Web Company Partners with WAND.” https://www.poolparty.biz/news-events/press-release-the-joint-offer-of-wands-taxonomies-and-poolparty-taxonomy-management-accelerates-client-time-to-delivery/.

Sharma, Rita, David Poole, and Clinton Smyth. 2010. “A Framework for Ontologically-Grounded Probabilistic Matching.” *International Journal of Approximate Reasoning* 51 (2): 240–62. https://doi.org/10.1016/j.ijar.2009.05.007.

Sowa, John. 2005. “Conceptual Graphs.” 2005. https://www.jfsowa.com/cg/index.htm.

Stanford University. 2022. “Protégé.” 2022. https://protege.stanford.edu/.

“Upper Ontology.” 2022. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Upper_ontology&oldid=1073911935.

W3C. 2012. “OWL - Semantic Web Standards.” 2012. https://www.w3.org/OWL/.

———. 2015. “Corine Land Cover Nomenclature in SKOS.” 2015. https://www.w3.org/2015/03/corine.

WikiData. 2022. “Wikidata.” 2022. https://www.wikidata.org/wiki/Wikidata:Main_Page.

Wikipedia. 2022. “Semantic Network.” In *Wikipedia*. Wikipedia. https://en.wikipedia.org/w/index.php?title=Semantic_network&oldid=1074130855.

Wüster, Wolfgang, Scott A Thomson, Mark O’shea, and Hinrich Kaiser. 2021. “Confronting Taxonomic Vandalism in Biology: Conscientious Community Self-Organization Can Preserve Nomenclatural Stability.” *Biological Journal of the Linnean Society* 133 (3): 645–70. https://doi.org/10.1093/biolinnean/blab009.

Yang, Guang, Qinghao Ye, and Jun Xia. 2022. “Unbox the Black-Box for the Medical Explainable AI via Multi-Modal and Multi-Centre Data Fusion: A Mini-Review, Two Showcases and Beyond.” *Information Fusion* 77 (January): 29–52. https://doi.org/10.1016/j.inffus.2021.07.016.

Yoon, Carol Kaesuk. 2009. *Naming Nature: The Clash Between Instinct and Science*. W. W. Norton & Company.

ABSTRACT IN FRENCH BELOW

TERMINOLOGIE, TRADUCTION ET INTELLIGENCE ARTIFICIELLE

RÉSUMÉ

Un grand nombre de noms ou de syntagmes nominaux utilisés par les humains peuvent être catégorisés selon des taxonomies dans lesquelles un nom ou un syntagme nominal désigne une chose qui est un type de chose à laquelle se rapporte un autre nom ou un autre syntagme nominal. Par exemple, une vache est un type d'animal. Il est essentiel de connaître ces liens taxonomiques pour un raisonnement juste, que celui-ci soit fait par un humain ou par un ordinateur. Malheureusement, les liens taxonomiques entre les noms et les syntagmes nominaux sont souvent confondus avec les liens méronymiques (partie d'un tout) ou les deux concepts sont confondus dans une structure hiérarchique, comme le permet le Système simple d'organisation des connaissances (SKOS) qui est très répandu.

Pour résoudre ce problème, les spécialistes de l'intelligence artificielle ont mis au point des applications qui permettent de faire une distinction entre les taxonomies et les méronymies. Ces applications permettent également d'exprimer les taxonomies sous forme de graphiques plutôt que sous forme de structures hiérarchiques et ainsi de tenir compte des propriétés multiples léguées au fil du raisonnement. Le déploiement réussi de ces applications repose sur une analyse minutieuse de la définition de chaque nom ou syntagme nominal dans une taxonomie. Par conséquent, dans le contexte de la traduction des langues, ces applications peuvent jouer un rôle important en révélant les différents sens des noms ou des syntagmes nominaux dans différentes langues qui devraient avoir la même signification. Si elles ne sont pas résolues, ces différences entraîneront un transfert des connaissances inadéquat dans les systèmes d'intelligence artificielle fonctionnant dans plus d'une langue, ainsi que des attributions et des explications incorrectes par ces systèmes.

Postal address for correspondence:

Clinton Smyth

1932 Charles St
Vancouver
British Columbia
Canada   V5L 2T9

Phone: + 1 604 812-2596

Email: cpsmyth@msn.com