Ontology Patterns Guidelines

version 1.0.2 · 8 May 2024

This document and its contents were created by Cesar Gonzalez-Perez, Martín Pereira-Fariña and Beatriz Calderón-Cerrato.

Partial funding was provided by Incipit · CSIC, USC, and project ACME (grant number PID2020-114758RB-I00 funded by MCIN/AEI/10.13039/501100011033, Retos de la Sociedad, Plan Estatal 2017-2020, Spain).

8 May 2024 14:34 · revision 1320

Table of Contents

Introduction

This document provides guidelines for the ontological analysis of discourses. The purpose of ontological analysis is to produce an ontology, that is, a model of the things in the world that the text refers to, their properties and their relationships.

This document is designed as a pattern reference, that is, it describes what to do when a certain pattern is found during ontological analysis. Guidelines are given in the form of situation/solution pairs, thus indicating what solution must be applied when a particular situation is found. Also, examples are used throughout.

Please see the *IAT/ML Analysis Process Guidelines* document for additional context and process-oriented guidance, and the *IAT/ML Technical Specification* document for specific details.

Organising Ontologies

This section provides guidelines related to the organisation of ontologies when working with multiple documents.

Ontology Trees

When doing ontological modelling of multiple documents throughout a corpus, it is a good idea to include different levels of detail at different levels in the corpus. For example, imagine a corpus containing a few topics, each topic containing some documents. The corpus ontology should contain only the entities that are common to the whole set of documents. Each topic should take this ontology and add details to it, but only those details that are present throughout all or most of the associated documents. Finally, each document should take its topic ontology and add the necessary details to describe specific entities that appear only in the document. In this manner, document ontologies derive from topic-level ontologies, which in turn derive from an overall corpus-level ontology, thus making up an ontology tree.

As a rule, atoms representing singular entities (such as *Adolf Hitler* or *Paris*) should be included in a corpus-level ontology if they appear repeatedly across the whole corpus. Similarly, they can be included in a topic-level ontology if they appear in most of the topic's documents. However, atoms that represent document-specific entities should be added to the corresponding document-level ontology only, not the topic or corpus-level ones.

Ontology Trees
Situation
Ontological modelling is to be performed for multiple documents in a corpus, perhaps organised in topics.
Solution
Organise ontologies as a tree over the corpus, its topics and documents, so that the ontology at each level contains entities that apply to all or most of its documents.

Reference Ontologies

When analysing documents that pertain to a particular domain, it may be the case that an ontology for that domain already exists. For example, the CHARM ontology (www.charminfo.org) is available for anything related to cultural heritage. Existing ontologies with the right scope can work very well as reference ontologies, that is, ontologies at the corpus or topic level from which document ontologies are derived by adding or changing the necessary details.

When defining a thematic corpus or a topic in a corpus, consider whether a reference ontology already exists, and consider adopting it if possible. Then, derive subtopic- or document-level ontologies from it. This saves work and improves interoperability.

Reference Ontologies
Situation
A thematic corpus or a topic within a corpus is being set up for ontological analysis.
Solution
Find out whether a reference ontology exists for the theme at hand, and adopt it if possible. Derive lower-rank ontologies from it.

Using Identifiers

It is often necessary to include representations of the same things in multiple ontologies. For example, each document that refers to the Eiffel Tower would contain an ontology having an atom representing the Eiffel Tower. When this happens, the problem arises of ensuring that the identity of the multiple elements representing the same thing is correctly expressed. Identifiers are used for this purpose.

When creating an ontology element that represents a thing that has already been represented in another ontology, make sure that you use the exact same identifier that was previously used. For example, the *Place* category in CHARM has the identifier affe312a-2cb4-40e8-a6dd-6eb3e7f51f27:174. This identifier was created when the *Place* category was defined in CHARM, and will never change. If you want to include the *Place* category in another ontology, make sure that you use the same identifier, so that the two categories are recognised as identical when integrating the ontologies or performing any operation that involves the two. This is valid even if the two categories have different names. For example, one ontology may call it “Place” and another could call it “Location”; as long as both categories use the same identifier, they will be considered to refer to exactly the same thing.

Using a reference ontology (see *Reference Ontologies*, page 4) is a good way to obtain and use well-established identifiers.

Using Identifiers
Situation
The same thing must be represented in two or more ontologies, and guarantees are needed that each of the ontology elements will be recognised as identical.
Solution
Create each ontology element with the same identifier. If you are using a reference ontology, use the identifiers provided by it in lower-rank ontologies.

Interpreting Text

This section provides guidelines related to the adequate interpretation of text so that faithful ontology elements can be obtained.

Prototypical Entities

Expressions that use a proper name or another kind of proper definite descriptor in singular, such as “Adolf Hitler”, “my house” or “the first car that my parents ever owned”, point at singular entities with a clear identity. However, expressions that use other kinds of descriptors, such as “the leafless trees” or “any visitor”, point at entities whose identity cannot be clearly established. For example, in the sentence “Any visitor that stays over one month has the right to use the phone”, “Any visitor” does not refer to any particular visitor, but to a prototypical entity. These entities can be modelled as atoms as if they were singular individuals, using a name that expresses this fact, such as *any_visitor*.

Prototypical Entities
Situation
A singular noun phrase referring to a prototypical entity must be modelled.
Solution
Model the entity as an atom, using an identifier that clearly indicates that it is not a singular entity.

Plurals

Plural noun phrases can be especially difficult to interpret in terms of identity. When a text uses a plural noun phrase, you must determine whether it is meant to describe a collection of specific entities, where the identity of each of them is relevant; or, to the contrary, whether the text aims to describe a generic collection of indeterminate entities or even a category. Consider the following examples:

A. “All my three children are very good at music”.

B. “Some children will inevitably knock on the door tonight”.

C. “Children are beautiful creatures”.

In example A, “All my three children” refers to very specific entities in the world, namely, my children. I even know their names and can visualise their faces. However, “Some children” in example B points to indeterminate children, with no particular names or faces. Furthermore, “Children” in example C does not point to a collection as such, but instead to the abstract notion of children, that is, the *Child* category. The identities of the entities referred to by A are crucial, as they cannot be replaced by others. However, identities are not relevant in examples B or C.

Plurals like in example A are called *identity plurals*, those like in example B are called *generic plural*s, and those like in example C are called *categorical plurals*.

For example, “All my three children” in example A should be modelled as a group of people containing three individuals, perhaps with their particular names and ages. “Some children” in example B could be modelled as a “Door-Knocking Child” category (or, perhaps, a “Door Knocker” role for the “Child” category; see *Identifying Roles*, page 19). Finally, “Children” in example C would be modelled as a “Child” category.

Plurals
Situation
A plural noun phrase must be modelled.
Solution
If the phrase is an identity plural, model it as a collection of atoms.
If the phrase is a generic or categorical plural, model it as a category or a role of a category.

Enumerations

An enumeration is a list of things, usually separated by “and”, “or”, or a similar connective. Enumerations may also appear as bulleted or numbered lists. Consider the following example:

> Books and magazines were catalogued and given a code.

Here, the text is enumerating “Books and magazines”. It is a list with only two items, but it is still a list. Consider another example:

> The riots were led by members of the party, shop owners, and a few bypassers.

Here, the text is enumerating “members of the party”, “shop owners”, and “a few bypassers”.

Enumerations can list verbs or adjectives in addition to nouns. Consider the following example:

> Suspects were identified, taken to the police station, and questioned.

Here, three verbs are being enumerated: “identified”, “taken to” and “questioned”.

In general, treat items in enumerations as if they appeared in isolation, and include each individual item in the appropriate terms list. In the case of nouns, introduce an abstract term for all the enumerated items if the text makes any statement that applies equally to all of them. In the first example, “Publication” should be introduced as an abstract noun in relation to “Books and magazines”, as both books and magazines were catalogued and given a code. Similarly, members of the party, shop owners, and a few bypassers in the second example share the fact that all of them led the riots, so “Riot Participant” or even “Person” should be added as an abstract noun.

Enumerations
Situation
An enumeration of nouns, verbs or adjectives must be modelled.
Solution
Treat each item in the enumeration as if it appeared in isolation.
In the case of noun enumerations, introduce an abstract noun if something is said in the text that applies equally to all the items in the enumeration.

Loaded Terms

Sometimes, terms in texts may be heavily loaded with political, religious, social, ideological, personal or otherwise sensitive meaning. If the text involves multiple speakers, it is possible that each one uses a different term to point to a common concept but with different connotations or semantic loads. Consider the following example:

> Alice: Many sex workers have chosen their job freely and we shouldn’t prohibit them from doing what they want.

> Bob: I disagree. All victims of prostitution must be protected and offered an alternative way of life.

Here, Alice is referring to “sex workers”, which connotes that prostitutes are genuine workers, whereas Bob uses the term “victims of prostitution”, which connotes the idea that prostitutes are victims. As an analyst, you should not take a side, and stay as detached as possible from the opinions, preferences and positions expressed by the speakers. In this regard, you should choose terms that are as non-committal and free from positioning as possible. In this example, “prostitute” would probably be better than either “sex worker” or “victim of prostitution”. In any case, this must be judged by each analyst depending on the topic and the context of the text.

Verbs
Situation
One or more speakers use heavily loaded terms, showing strong political, religious, social, ideological, personal or otherwise sensitive connotations.
Solution
Reword the terms so that they stay as non-committal and opinion-free as possible.

Creating Term Lists

This section provides guidelines related to the creation of term lists for nouns, adjectives and verbs.

Selecting Terms

A large piece of text may contain many terms. You don't want to include all of them in your term lists. Rather, you should include only those that are central to what the text is about. A typical ontology contains between 10 and 30 entities; if you have more than 50 terms, consider trimming down the lists and discarding the less relevant ones. You can always come back to the text later on and add missing pieces if you find the need. An ontology must be minimalistic.

When looking for relevant terms, consider different kinds of things that the text may refer to, including:

- Physical things such as objects or buildings
- Places
- People, including groups and individuals
- Expressive elements such as songs, performances or artworks
- Abstract ideas such as opinions or values
- Events
- Situations and circumstances
- Measurements
- Time spans such as moments or periods

The following sections provide advice on how to treat nouns, adjectives, verbs and adverbs.

Selecting Terms
Situation
You must select what terms to include in your lists.
Solution
Select only terms that refer to central elements in the text, and discard the rest. Focus on different kinds of thing that the text may refer to, such as physical things, places, people, expressive elements, abstract ideas, events, situations and circumstances, measurements or time spans.

Atom Nouns

Atom nouns are those that denote individual things, or atoms, such as "the Eiffel Tower" or "the mayor of London", as opposed to nouns that denote types of things, or categories, such as "people" or "buildings". Consider the following example:

> They spent a few days between York and another town ten miles to the North.

Here, the text is mentioning two towns: York and "another town ten miles to the North". Atom nouns must be added to the list as they appear, making sure that the specific thing in the world that they denote is perfectly recognisable even when read without context. "York" is clear as is, but we do not know the name of the second town, so we add the complete expression to the noun list, rewording it as necessary to make it as self-contained as possible:

- A town ten miles to the North of York

- York

Atom Nouns
Situation
An atom noun is found in the text and must be incorporated to the atom noun list.
Solution
Add the atom noun as it appears, changing it as little as possible. Ensure that it is understandable as added to the list, even when read without context. Reword the noun slightly if necessary for this purpose.

Category Nouns

Category nouns are those that denote types of things, or categories, such as “people” or “buildings”, as opposed to nouns that denote individual things, or atoms, such as “the Eiffel Tower” or “Bill Gates”. Consider the following example:

> Prisoners of war, especially women, were victims of gender-based violence across different countries.

Here, the text includes clear category nouns such as “prisoners of war”, “women” and “countries”. All of these are countable nouns, that is, we can enumerate their instances, as in “one woman, two women, three women”. Items from countable nouns must be added to the category noun list in singular and with initial capitals, like this:

- Country
- Prisoner of War
- Woman

In addition, the text mentions “gender-based violence”. However, “violence” is an uncountable noun, that is, we cannot possibly enumerate “one violence, two violences”. For this reason, we need to rephrase this noun in order to make it countable. One option would be “violence event”, that is, an event where violence is committed. This is a countable noun, so it can be added to the list:

- Gender-Based Violence Event

In addition, some nouns may only have a plural form. Some examples are “remains” or “belongings”, which cannot be put in singular. For nouns like these, you must also rephrase them, or find synonyms, in order to make them countable and put in singular.

Category Nouns
Situation
A category noun is found in the text and must be incorporated to the category noun list.
Solution
If the noun is countable, add it in singular and initial capitals. If it is uncountable or does not admit a singular form, rephrase it to make it countable, and add it similarly.

Adjectives

Adjectives are words or clauses that qualify nouns, either atom or category ones. Consider the following example:

> Many elderly patients did not survive the pandemic, and those who had a pre-existing respiratory condition were high-risk too. The most experienced ER doctors were called in to the hospital.

Here, both “elderly” and “those who had a pre-existing respiratory condition” are qualifying “patients”, whereas “the most experienced” and “ER” are qualifying “doctors”. Adjectives must be added to the list in singular, positive degree, present tense, with initial capitals, and accompanied by the atom or category noun to which they apply in parentheses:

- Elderly (Patient)
- ER (Doctor)
- Experienced (Doctor)
- Having a Pre-existing Respiratory Condition (Patient)

Notice that items in the adjective list may be actual adjectives but also full relative clauses (such as “who had a pre-existing respiratory condition”) or even nouns working as adjectives (such as “ER” in our example).

Adjectives
Situation
An adjective is found in the text and must be incorporated to the adjective list.
Solution
Add the adjective in singular, positive degree, present tense, with initial capitals, and accompanied by the atom or category noun to which they apply in parentheses.

Verbs

Verbs are words or clauses that indicate states, changes or actions.

Verbs can be classified according to their valency:

- **Impersonal** verbs have no subject or objects, such as in “It’s raining”. Impersonal verbs are rare, and you can usually ignore them.
- **Intransitive** verbs have a subject but no objects, such as in “Alice was dreaming”.
- **Transitive** verbs have a subject plus an object, as in “Bob bought a house”.
- **Ditransitive** verbs have a subject and two objects, as in “Claire gave Alice a book”.
- **Higher-valency** verbs have a subject plus three or more objects. These are very rare, and they are often easy to rephrase as simpler constructions.

Also, bear in mind that verbs that indicate more or less stable states are easier to represent in an ontology than verbs indicating events or rapid changes, and they usually provide more value. Consider the following example:

> Museum staff packaged the objects in boxes, and then the boxes were sent to different museums.

Here, “packaged” and “were sent to” are verbs that describe events. Often, verbs like these can be easily rephrased in such a way that they express states. For example:

- “Museum staff packaged the objects in boxes” can be rephrased as “Boxes contain objects” or “Objects are located in boxes”.
- “boxes were sent to different museums” can be rephrased as “boxes are located in museums”.

The following sections provide specific advice for verbs of different valency.

Verbs
Situation
A verb is found in the text and must be incorporated to the verb list.
Solution
Determine the valency of the verb and follow the corresponding advice. If the verb represents a rapid change or event, and can be easily reworded to indicate a state, do it. In any case, verbs must be added in the present tense and with initial capitals.

Intransitive Verbs

Intransitive verbs are those that have a subject but lack an object, such as in "Alice dreams often".

Intransitive verbs should be rephrased as adjectives by using an *"is" + gerund* pattern. For example, "Alice dreams often" can be dealt with by adding "Alice" to atom noun list plus "Is Dreaming (Alice)" to the adjective list (see *Adjectives*, page 10).

Intransitive Verbs
Situation
An intransitive verb is found in the text and must be incorporated to the verb list.
Solution
Add the verb subject to the corresponding noun list, and the verb itself, following an *"is" + gerund* pattern, to the adjective list.

Transitive Verbs

Transitive verbs are those that have a subject plus an object, such as in "Bob bought a house". Consider the following example:

> Many elderly patients did not survive the pandemic, and those who had a pre-existing respiratory condition were high-risk too.

Here, "did not survive" is a transitive verb, and should be added to the verb list. Words like "had" in "had a pre-existing respiratory condition", or "were" in "were high-risk", are verbs too, but they should not be added to the list, for the following reasons. In the case of "had a pre-existing respiratory condition", this is better modelled as an adjective (see *Adjectives*, page 10). In the case of "were high-risk", the web *to be* is a special case that requires special treatment (see *Handling "To Be"*, page 16). Verbs added to the list must be in present tense, with initial capitals, and accompanied by the atom or category nouns that work as their subject and object in parentheses:

- (Patient) Survives (the pandemic)

Transitive Verbs
Situation
A transitive verb is found in the text and must be incorporated to the verb list.
Solution
Add the verb to the verb list, accompanied by the atom or category nouns that work as subject and object, in parentheses.

Ditransitive Verbs

Ditransitive verbs are those that have a subject and two objects, such as in "Claire gave Alice a book".

Ditransitive verbs should be recast as nouns expressing a situation, process or event. In the previous example, there is a giving event involving Claire as the giver, Alice as the recipient, and the book as the thing being given. Notice that now we have four nouns and three transitive verbs. Add the four nouns to the corresponding noun list, and each of the three transitive verbs to the verb list by following the advice in *Transitive Verbs*, p. 12. In this example, the atom noun list would look as follows:

- a book
- Alice
- Claire

The category noun list would look as follows:

- Giving Event

And the verb list would look as follows:

- (Alice) Gives In (Giving Event)
- (Book) Is Given In (Giving Event)
- (Claire) Receives In (Giving Event)

Ditransitive Verbs
Situation
A ditransitive verb is found in the text and must be incorporated to the verb list.
Solution
Rephrase the ditransitive verb as a noun expressing a situation, process or event. Add it to the noun list, together with the subject and objects. Add the three separate transitive verbs connecting the four nouns to the verb list.

Adverbs

Adverbs are words or clauses that qualify verbs or adjectives. Consider the following example:

> Mining infrastructures are usually built as rapidly as possible in order to minimise the impact of changes in legislation.

Here, "as rapidly as possible" is qualifying "built", which is a verb. In order to incorporate an adverb to your lists, you must first rephrase the verb as a situation, process or event, that constitutes a noun. In our example, you can recast "built" as a "building event" or "construction event":

- Construction Event

Then, you can add the adverb to the adjective list as an adjective:

- Rapid (Construction Event)

This kind of reformulation of a verb as a noun is similar to the process described to deal with ditransitive verbs (*Ditransitive Verbs*, p. 13).

If the adverb is modifying an adjective, then make it part of the adjective itself. Consider the following text:

> The load was very heavy, which made the structure collapse.

Here, “very” is an adverb qualifying the adjective “heavy”. In cases like this, add the “very heavy” adjective to the adjective list as a single item.

Adverbs
Situation
An adverb is found in the text and must be incorporated to the appropriate term list.
Solution
If the adverb is qualifying a verb, rephrase the verb as a noun expressing a situation, process or event, and add it to the noun list, together with the subject and objects. Add the three separate transitive verbs connecting the four nouns to the verb list. Finally, add the adverb as an adjective to the adjective list. If the adverb is qualifying an adjective, add the adjective plus adverb to the adjective list as a single item.

Processing Term Lists

This section provides guidelines related to the processing, checking and simplification of term lists.

Removing Redundancy

It is quite likely that the text you are analysing uses different terms to refer to the same thing, or that you have introduced duplicate or otherwise redundant items in term lists. Consider the following example:

> Most mining sites require a license from the local government before they can operate. In order for a site to get licensed, owners need to apply at least three months in advance.

Here, the category nouns “mining sites” and “sites” are most likely redundant, as they refer to the same category of things. You may think that mining sites are a subtype of sites, but since the text does not mention any other kind of site, then we can safely assume that all the sites the text is concerned with are mining sites, so we can conflate both nouns. To make sure, try replacing on term with the other (“mining site” with just “site”, in our example) throughout the text and see if it makes sense. Try also the inverse replacement. If both work, it is likely that the terms are redundant. If they are, then we must remove one and leave the other. In our example, we should decide whether we want to use “Site” or “Mining Site”, and remove the other one from the list.

Redundancy may also occur across lists. Consider the following example:

> Some mining sites are developed in an open-pit manner, on the surface. Open-pit sites have a large environmental impact and require special treatment.

After preliminary analysis, our adjective list contains an item for “Open-Pit (Site)”, whereas our verb list contains an item for “(Site) Is Developed In (Open-Pit Manner)”. Obviously, this adjective and this verb refer to the same thing, so we should remove one.

When redundancy happens across lists, try to favour keeping a noun over an adjective, and an adjective over a verb. In our example, we should keep the “Open-Pit (Site)” adjective and remove the verb.

Removing Redundancy
Situation
Two terms (in the same list or different lists) look like they may refer to the same thing.
Solution
Examine the text carefully to determine whether one term can be replaced for the other meaningfully. Repeat inverting the term replacement. If the text still makes sense, consider removing one of the terms. Favour shorter and simpler terms over longer ones. Favour keeping nouns over adjectives, and adjectives over verbs.

Checking Integrity

The four terms lists are connected, as items in the adjective and verb lists contain references to items in the noun lists. Also, items in the atom noun list contain references to items in the category list. References within lists also exist. For example, items in the noun lists representing properties or roles contain references to other items in the same list.

You must make sure that these references work well. In particular, ensure that every reference points to an item in a list. For example, if you have the verb "(Doctor) Is Called in To (Hospital)", make sure that both "Doctor" and "Hospital" exist in the category noun list. If either does not exist, then the reference is broken, and you should fix it:

- Start by trying to find the appropriate noun in the atom or category noun list. For example, if "Doctor" is not in the category noun list, perhaps there is a "Health Professional" item. If you think this is what "Doctor" means, then fix the broken reference by changing "Doctor" to "Health Professional" throughout.
- If you can't find any suitable item to fix the reference, you must add it. In our example, you may need to add "Doctor" to the category noun list if there is no appropriate term for it.

In any case, you must end up with a fully connected mesh of list items that contains no broken references.

Checking and Fixing Integrity
Situation
An item in the adjective or verb list makes a reference to a noun that does not exist in either of the noun lists.
Solution
Find an item in one of the noun lists that can replace the broken reference. If there is none, add it.

Handling "To Be"

The verb "to be" in English (and equivalent verbs in other languages) often works as a copula, that is, it appears by itself rather than as an auxiliary in a larger verbal construction. For example, "is" in "a mining site is usually composed of multiple pits" plays an auxiliary role, as the complete verb here is "To Be Composed Of". This makes a perfectly acceptable item in your verb list and needs no special treatment. However, "is" in "Red Hill is a mining site" is not auxiliary but a copula, so it does need special treatment as described in this section. When working as a copula, "to be" may mean many things. It can mean existence ("there is a mining site"), identity ("Red Hill is the oldest mining site in the country"), predication ("Red Hill is abandoned"), classification ("Red Hill is a mining site") or subsumption ("mining sites are industrial facilities").

It is likely that your adjective or verb lists will contain instances of the verb "to be" as a copula. For each one, determine whether it means existence, identity, predication, classification or subsumption, and act accordingly:

- **Existence** instances can be safely removed, because the existence of something is already contemplated by that thing being present in a noun list. For example, if the atom noun list contains "Red Hill", then you do not need "there is a mining site named Red Hill" in the verb list.
- **Identity** instances mean that there are two nouns referring to the same thing. For example, "Red Hill is the oldest mining site in the country" implies that "Red Hill" and "the oldest mining site in the country" refer to the same thing. For this reason, you should remove the item from the verb list, drop one of the two nouns, and convert the other in a feature or facet of the former. For example, you could keep "Red Hill" and remove "the oldest mining site in the country", and capture the fact that Red Hill is the oldest mining site in the country by adding "Oldest (Red Hill)" to the adjective list.

- **Predication** instances should be removed from the verb list and replaced by an item in the adjective list. For example, “Red Hill is abandoned” would be replaced by “Abandoned (Red Hill)”.
- **Classification** instances should be removed from the verb list and replaced with a pair of items, one in the category noun list and another one on either noun list. For example, “Red Hill is a mining site” should be replaced by “Mining Site” or simply “Site” in the category noun list plus “Red Hill” in the atom noun list.
- **Subsumption** instances should be removed from the verb list and replaced with two items in the category noun list, indicating the subtyping relationship. For example, “mining sites are industrial facilities” should be removed by “Industrial Facility” plus “Mining Site (subtype of Industrial Facility)” in the category noun list.

In this manner, instances of “to be” as a copula should disappear from the verb list and be replaced by items in the nouns or adjective lists.

Handling “To Be”
Situation
An instance of the verb “to be” as a copula appears in an item of the verb list.
Solution
Determine whether “to be” indicates existence, identity, predication, classification or subsumption. Remove the item from the verb list and add any necessary noun or adjective items, depending on the case.

Fixing Vague Language

Sometimes, the terms used in the text being analysed may be too vague as to convey a clear and strong meaning. Consider the following example:

> Ancient settlements are often related to rivers or other bodies of water. Among the many things often found in sites, pottery fragments are very common.

From this text, we would be tempted to add a “(Settlement) Is Related To (Water Body)” item to the verb list. However, “being related to” is a very vague statement. What does the text mean? Similarly, “things” is too vague, and we should not add a “Thing” item to the category noun list.

Any term that is too vague must be revised and replaced by a stronger one. The text may provide context to help with finding a good replacement, but sometimes you may need to check with external sources or a domain specialist. In the example above, “Is Related To” probably means “Is Close To”, and “Thing” probably means “Archaeological Artefact”.

Fixing Vague Language
Situation
A term in a list is too vague or weak, so its meaning is not clear.
Solution
Replace the vague term with a stronger one. Find context in the text that helps you decide what term to use. If necessary, look for additional information in external sources or consult with domain experts.

Identifying Noun Characteristics

Some items in the noun lists may be features or facets rather than proper entities. Consider the following example:

> Every object donated to the museum was given a code and a description by staff, and stored in a box.

The category noun list for this text would include the following items:

- Box
- Code
- Description
- Member of Staff
- Object

Here, it is clear that “Code” and “Description” do not constitute entities of themselves, but are instead features of “Object”. We must indicate this in the list by adding the entity they refer to in parentheses:

- Box
- Code (of Object)
- Description (of Object)
- Member of Staff
- Object

In addition, the verb list would contain the following item:

- (Object) Is Stored In (Box)

Similar situations may occur in relation to atom rather than category nouns. Consider the following example:

> Red Hill is located in northern New South Wales and occupies an area of over 24 square kilometres.

The atom noun list for this text would include the following items:

- Northern New South Wales
- Red Hill

And the adjective list would contain the following item:

- area of 24 square kilometres (Red Hill)

In addition, the verb list would include the following item:

- (Red Hill) is located in (Northern New South Wales)

Note that we use the word “features” for category items and “facets” for atom items.

Identifying Noun Characteristics
Situation
An item in a noun list constitutes a characteristic of another item, rather than an entity of itself.
Solution
If the item is relating two items together, then add it to the verb list. If the item is not relating two items together, but simply characterising one, add the item that is being characterised in parenthesis.

Adding Types for Instances

The atom names list, by definition, contains items that refer to individual things. In order to construct a proper ontology, it is highly recommended that every atom can be classified under

a particular category, so you must make sure that the category names list contains items for every item in the atom noun list. Consider the following example:

> The Red Hill mining site was one of the first in the region, and the largest. It was created and owned by the New South Wales Mining Corporation until 2012.

From this text, the atoms noun list would contain items "Red Hill" and "New South Wales Mining Corporation". The first one is a mining site, so we should make sure that the category noun list contains a "Mining Site" item. In turn, "New South Wales Mining Corporation" is a company, so the category noun list should contain a "Company" item.

Similarly, the adjective list may contain a "Large (Red Hill)" item. As before, we must ensure that a type-level item exists for this instance-level one. Since the adjective "large" refers to size, we could expand "Mining Site" in the category noun list as "Mining Site (features: Size)".

Finally, it is likely that the verb list contains "(Red Hill) Is Created By (New South Wales Mining Corporation)" and "(Red Hill) Is Owned By (New South Wales Mining Corporation)", as these two facts are clearly stated by the text. As for the nouns, we should make sure that type-level items exist in the verb list for these two instance-level items. For example, we could add "(Mining Site) Is Created By (Company)" and "(Mining Site) Is Owned By (Company)".

As you identify and add type-level items, annotate each instance-level item with its type, in parentheses, as follows:

- New South Wales Mining Corporation (instance of Company)
- Red Hill (instance of Mining Site)

In general, make sure that no instance-level item in your lists lacks a corresponding type-level item. All the items in the atom noun list, as well as the instance-level items in the adjectives and verb list, should have an indication of their type in parenthesis.

Adding Types for Instances
Situation
An instance-level item appears in a terms list.
Solution
Make sure that there is a category-level item in a list to which the instance-level item corresponds. If there is none, add it. For nouns, add it to the category noun list. For adjectives, add them to the category noun list. For verbs, add it to the verb lists. Annotate each instance-level item with a reference to its type-level one in parentheses.

Identifying Roles

A role is a particular point of view that we take on a category or atom in the context of an association or link. Consider the following example:

> University staff that enrol as students in undergraduate courses will get a 20% discount.

From this text, our category noun list would probably contain the following items:

- Student
- Undergraduate Course
- University Staff Member

However, University Staff Member and Student are not proper categories but roles that people play in the context of their relationship to a university. Roles can be identified from the following clues:

- **Simultaneity**. Some things may be categorised as two or more roles at the same time. For example, and as expressed by the example above, one person can be both a university staff member and a student simultaneously.
- **Temporality**. The things that play the roles can start and stop playing them over time without losing its identity. In our example, someone can become a student or a staff member, or cease being either of them, without losing their identity as a person.
- **Type constraint**. Playing the role is restricted to things of a particular kind. In our example, only atoms of the Person category may be students or university staff members. Dogs or trees cannot play these roles.

If a category noun has these properties, then it is likely a role rather than a category. In this case, determine the type constraint category, Add it to the list if not there, and mention it as a role of the necessary categories, in parentheses. In our example, the list would end up as follows:

- Person
- Student (role of Person)
- Undergraduate Course
- University Staff Member (role of Person)

Identifying Roles
Situation
One item in the category noun list is a role of another.
Solution
Add the category playing the role to the list if not already present. Add it to the constraint type category item in parentheses.

Identifying Supertypes

It is common that texts include terms that are connected by subtyping or supertyping relationships, sometimes referred to as subsumption. Consider the following example:

> Ancient settlements are often close to rivers, lakes or the ocean. Among the many types of remains often found in sites, pottery fragments and crumbling walls are perhaps the most common.

From this text, our category noun list would probably contain the following items:

- Lake
- Ocean
- Pottery Fragment
- Remains Item
- River
- Settlement
- Wall

It is clear from the text that pottery fragments and walls are types of remains items. We must add this information to the list by writing the supertype of each item in parentheses, like this:

- Lake
- Ocean
- Pottery Fragment (subtype of Remains Item)
- Remains Item
- River
- Settlement

- Wall (subtype of Remains Item)

In addition, it is evident from the text that rivers, lakes and oceans share something in common: all of them are usually close to ancient settlements, and we know that all of them are bodies of water. Although the text simply enumerates them and does not provide an abstract noun that encompasses all of them, we should add it (see *Enumerations*, page 7). Once we have added it, we can mark its subtypes in the same manner as above:

- Lake (subtype of Water Body)
- Ocean (subtype of Water Body)
- Pottery Fragment (subtype of Remains Item)
- Remains Item
- River (subtype of Water Body)
- Settlement
- Wall (subtype of Remains Item)
- Water Body

Identifying Supertypes
Situation
One item in the category noun list is a subtype of another.
Solution
Indicate the supertype of the item in parentheses.

Constructing Ontology Elements

This section provides guidelines related to the definition of ontology elements from the term lists.

Creating Categories and Features

A category must be created in the ontology for each item in the category noun list. If there is a supertype assigned, make the category a subtype of that. If there are features associated to the item, decide, for each one, what type of feature it should be: a property or an association. Properties hold quantities or qualities, such as *Age* or *Name*, whereas associations point to other elements in the ontology, such as *Is Located In*, which points to a place.

In addition, if there are any roles associated to the item, consider whether they are satisfied by existing associations or whether you need to add new associations. For example, "Student (role of Person)" may imply that you need to associate *Person* to *University* via a *Studies In* association, in the context of which *Person* plays the student role.

Also, we must look at the adjective list, as it may contain additional properties for the category, such as "Decorated (Object)".

Finally, we must look at the verb list and find those items for which the category being created appears as subject or object. These items will constitute associations that must be added to the ontology as well. Define the associations in both directions. For example, if you find "(Monument) Is Located In (City)" in the list, create an association from *Monument* towards *City*, and another association from *City* towards *Monument*.

Creating Categories and Features
Situation
Categories and features are to be added to the ontology.
Solution
Create a category for each item in the category noun list. If the item has an associated supertype, make the category a subtype of it. If the item has associated features, create the necessary properties or associations, depending on whether the feature can hold quantities or qualities, or point to other elements in the ontology. Add extra associations, if needed, for any roles of the category item. Find the category being created in the adjective list, and create additional properties for any found items. Find the category being created in the verb list, either as subject or object, and create additional associations for any found items. Define the associations in both directions.

Creating Atoms and Facets

An atom must be created in the ontology for each item in the atom noun list. Make the atom an instance of the category that is associated to it.

Also, we must look at the adjective list, as it may contain values for the atom, such as "Is 300 metres high (the Eiffel tower)". Make values instances of the corresponding properties.

Finally, we must look at the verb list and find those items for which the atom being created appears as subject or object. These items will constitute references that must be added to the ontology as well. Define the references in both directions. For example, if you find "(the Eiffel

tower) Is Located In (Paris)" in the list, create a reference from *the Eiffel tower* towards *Paris*, and another reference from *Paris* towards *the Eiffel tower*. Make references instances of the corresponding associations.

Creating Atoms and Facets
Situation
Atoms and facets are to be added to the ontology.
Solution
Create an atom for each item in the atom noun list. Make it an instance of the associated category. Find the atom being created in the adjective list, and create additional values for any found items. Make values instances of the corresponding properties. Find the atom being created in the verb list, either as subject or object, and create additional references for any found items. Define the references in both directions. Make references instances of the corresponding associations.