Datamaran
Solution to the Human Rights ontology task
Alex García Laguía

Task 1: Critically revise the list of terms given to represent the topic

The list includes 79 terms that are related to the semantic fields of human rights and that intend to represent the field. The list has a flat structure and the elements appearing in the list are not related between them. The first thing we did was to organize these terms in groups and see how they relate to each other:

Term	Variation/s	Pattern/s
asylum_applicant	asylum_applicants	singular-plural
asylum_refugee	asylum_refugees	singular-plural
asylum_seeker	asylum_seekers	singular-plural
asylum_victim	asylum_victims	singular-plural
basic rights		
civic rights		
civil right	civil rights	
customary rights		
economic right	economic rights	singular-plural
fundamental rights		
human rights		
right of self-determination	right of self determination	hyphen / no hyphen
	right to self determination	right of / right to
	right to self-determination	prep + hyphen
right to be different		
right to education		
right to freedom of movement		
right to freedom of religion		
right to freedom of thought		
right to own property		
right to self-determination		
right to work		
access to potable water		
access to sanitation		
access to medical care	healthcare accessible	synonym + adj
freedom of choice		
freedom of information		
fundamental freedoms		
discrimination_abuse	discriminations_abuse	singular-plural
discrimination_color	discriminations_color	singular-plural and UK - US English
discrimination_colour	discriminations_colour	singular-plural
discrimination_equal	discriminations_equal	singular-plural
discrimination_equality	discriminations_equality	singular-plural and N - Adj
discrimination_origin	discriminations_origin	singular-plural
discrimination_race	discriminations_race	singular-plural
discrimination_racial	discriminations_racial	singular-plural and N - Adj
discrimination_religious	discriminations_religious	singular-plural
discrimination_rights	discriminations_rights	singular-plural
discrimination_sex	discriminations_sex	singular-plural
discrimination_sexist	discriminations_sexist	singular-plural and N - Adj
discrimination_sexual	discriminations_sexual	singular-plural and N - Adj
discrimination_social	discriminations_social	singular-plural
discrimination_society	discriminations_society	singular-plural and N - Adj
discrimination_vulnerable	discriminations_vulnerable	singular-plural
discriminatory practices		N - ADJ
nondiscrimination		Negation / antonymy
inhumane treatment		
racism	racial prejudice	N / Adj
	racial segregation	
protect respect		
un guiding principles		
ungp		

Table 1: grouping of the terms provided in the list

Superclasses and semantic relationships

- Most of the terms appearing can be semantically classified into two main groups, one group includes rights and the other includes discrimination, these two groups can be viewed as two main categories or superclasses: **rights and discrimination.**
- The two categories are **opposed semantically** and can be considered as some kind of **antonyms**. The rights appearing in the list are all human rights, which are by definition 'inherent to all human beings regardless of race, sex, nationality, ethnicity, language, religion (see https://www.un.org/en/sections/issues-depth/human-rights/). This definition is based on one of the core principles of human rights: the principle of non-discrimination, which guarantees that "human rights are exercised without discrimination". Thus, any kind of discrimination is opposed to this core caracteristic of human rigths and for this reason, the two superclasses can be viewed as being opposed to each other.

Subclasses and semantic relationships

- A number of more specific terms can be included in the superclass 'rights', as for example 'basic rights', 'civic rights', 'economic rights', 'fundamental rights' 'customary rights'. These terms may constitute subclasses of the 'human rights' superclass and each of them may include one or more of the rights appearing in the list. Some of these subclasses coincide with the traditional 5-class system of categorization of human rights: civil, political, economical, social and cultural rights.
- The category 'rights' may also include other subcategories provided in the list such as 'asylum', 'freedom' or 'access'
- The semantic relation between the category 'rights' and its subcategories is a relation of **hyponymy**, which is defined as a "relationship between a generic term (hypernym) and a specific instance of it (hyponym)".
- The terms 'asylum', 'freedom' and 'access' subcategories of human rights, and thus are **hyponyms** of the word 'human rights'. Conversely the term 'human rights' is a **hypernym** of the words 'asylum', 'freedom' and 'access'
- A third subclass of the superclass 'human rights' may be distinguished as well although the relationship between the two is not hyponymical. The category includes the terms 'protect respect' 'un guiding principles', 'ungp'. These terms refer to documents in which human rights are described in the context of business (UN Protect, Respect and Remedy" Framework), and provide a framework for their application as well as guidelines for their implementation (UN Guiding Principles on Business and Human Rights). A possible label for this category can be 'documents''.
- The different types of discriminations (colour, origin, race, religion, …) can also form subcategories and be viewed to be hyponyms of discrimation

Subsubcategories and semantic relationships

These subsubclasses include terms that are hyponyms of the terms we have labelled as subclasses, which in turn are hyponyms of the terms we posited as main categories. For

example the term 'right to education' is a subclass/hyponym of the 'civic rights', and civic rights is a subclass/hyponym of the term 'human rights'.

The three levels of categorization and their semantic relationships can be represented in the form of a taxonomy (note that not all subclasses are represented):
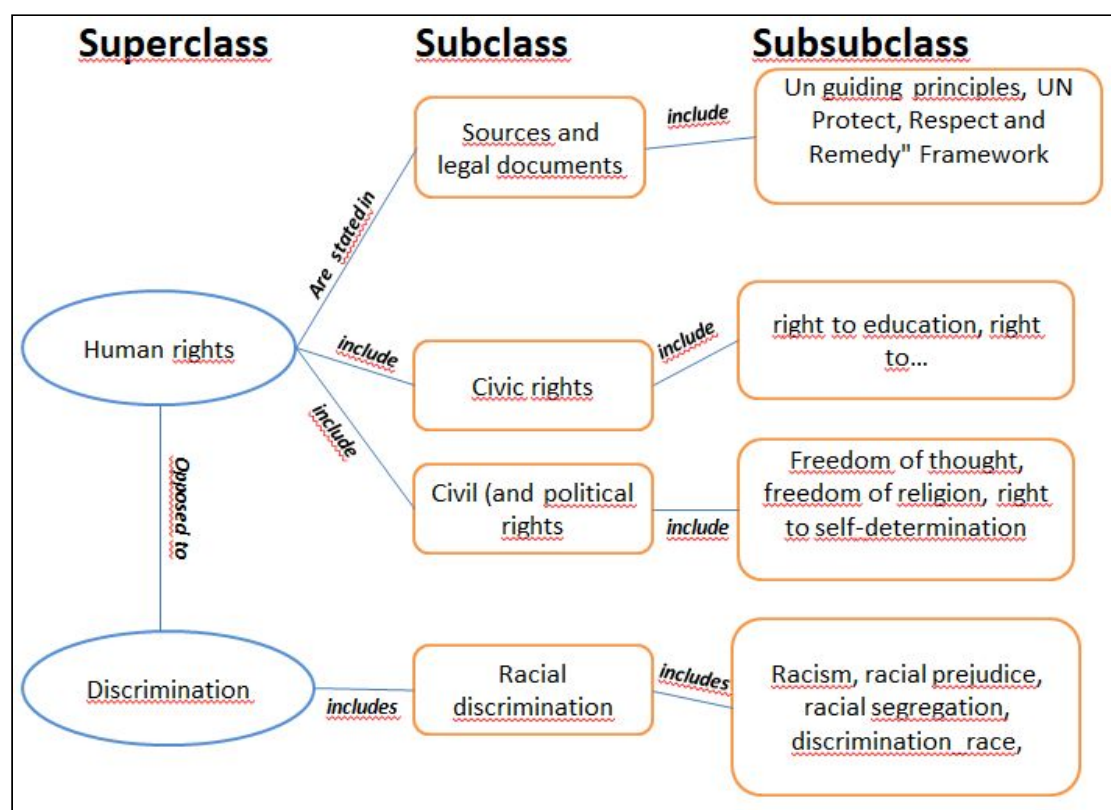

Table 2: Taxonomical organization of the terms provided in the list

Morphological relationships between the terms provided in the list.
As it can be seen in the Variation column of Table 1, a number of terms in the list can be related through morphological relationships rather than semantical.

- Most examples included in the superclass 'discrimination' include their plural equivalent 'discriminations'
- The terms included in the subclass 'asylum' also show their plural equivalent 'asylum_applicant' vs 'asylum_applicants'
- Other terms such as 'civil right' may also show the plural 'civil rights'
- Another type of morphological relationship between certain terms is a derivational relationship (which involves a change in the word class):
  - Noun to adjective
    - racism → racial
    - access → accessible
    - discrimination → discriminatory
- Some terms are spelled with a hyphen and also without it.

- ○ self determination / self-determination
  - The preposition following the word 'right' is also changed in certain cases.
    - ○ right to self-determination
    - ○ right of self-determination

Task 2: find ways to improve it, and, in the light of your revision, propose and make the case for considering a new ontology for the topic.

As we have seen above the taxonomical classification appearing in Table 2 can be considered as an improvement of the list provided for this task, given that it organizes a flat list of terms into classes and subclasses, and also defines the relations (semantic and morphological) among the different classes and subclasses.

An improved ontology[1] of human rights would:

1. have a purpose and a scope clearly defined. In the case of the list provided for this task it would seem as if the scope of the ontology should be limited to three main superclasses: 'rights', 'violations' and 'documents'
2. include an improved definition of human rights that would highlight the universality of this term. As we have seen above, some of the core principles on which human rights are based are useful for the definition of classes and subclasses of the taxonomy. An improvement on the definition provided for this exercise could be: *"references to internationally shared fundamental rights ensuring basic conditions for __all__ individuals to live in dignity"*
3. consider the existing literature as well as the opinion of experts in the topic of human rights. As it will be shown below, and assuming that the scope of the taxonomy is as wide as possible, we propose a reviewed taxonomy of human rights based on the following three documents:
   a. the Universal Declaration of Human Rights
   b. the International Covenant of Economic, Social and Cultural Rights
   c. the International Covenant of Civil and Political Rights
4. should consider additional subclasses such as entitlements
5. consider existing ontologies
6. be more consistent from a morphological perspective (as we saw above some terms show their plural equivalent but this is not the case for all terms on the list
7. be machine readable

---

[1] We follow the convention provided by Bermejo, J (2007, Simplified Guide to Create an Ontology), in which taxonomy and ontology are used as quasi-synonyms. According to the author, a taxonomy, which may also be called a lighweight ontology, does not include axioms or functions

8. be tested a number of times and accordingly improved, given that ontology development is regarded as iterative process in the specific literature.

Towards a new taxonomy

In relation with improvements suggested in 1 and 2 we provide the following upper nodes of a revised taxonomy, in which the category 'Human rights' is opposed to a category we have labelled as "Violations of human rights, that we consider to be more general and representative than the term 'discrimination'.
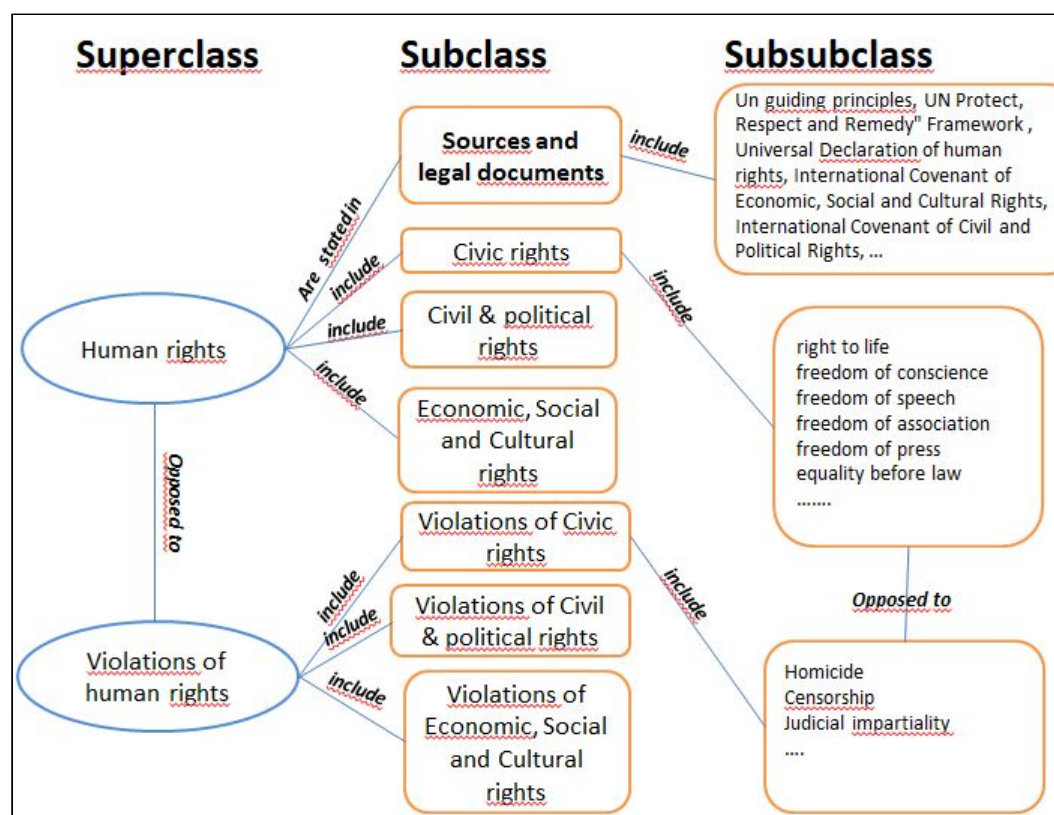

Table 3: revised taxonomy of human rights (higher nodes)

In addition, we propose to replace the subclasses of human rights by using the categories used by the United Nations in the covenants cited in improvement 3. The three resulting categories (Civic Rights, Civil & Political rights, Economic, Social and Cultural rights) seem to be used broadly in the literature related to the topics of human rights, as well as on the internet. Thus, in the same way as the two superclasses are opposed, each of the subclasses of human rights has a correspondent violation.

On the basis of these documents, we can also propose a revised taxonomy for each of the subclasses of Human rights. The taxonomy includes many of the terms

5

proposed in the original list of terms, and adds a number of new terms extracted from the Universal HR declaration as well as from the two covenants:

Individual / Civic rights	Civil rights	Political rights
right to life	right to life	freedom to think
right to liberty	right to be free from torture	freedom to believe in the concept's of one's religion
right to family	right to be free from degrading treatment	freedom to partake in the practices of one's religion
freedom of conscience	right to personal liberty, privacy,	righ to vote
freedom of speech	freedin of thought	right to political participation
freedom of association	freedom of religion	right to participate in public affairs
freedom of press	freedom of expression	freedom of expression
equality before law	freedom of information	freedom of assembly
right of law	freedomf of momevent with one's country	freedom of association
		right of self determination
		right to life

Cultural rights	Economic rights	Social rights
right to take part un cultural life	right to decent work	right to an adequate standard of living
availability of cultural goods	right to equal pay for work of equal value	adequate food
accessibility to culture	right to safe and healthy working conditions	adequate clothing
acceptability of the agents involved	right to a fair wage	adequate housing
adaptability of strategies, policies, pogreammes and measures	right to form trade unions	right to the continous improvement of living conditions
appropriateness	right of trade unions to establish national federations or confederations	right to educations
right to enjoy benefits of scientific progress and its applications	right of federations to join international trade union organizations	right to health
right to benefit from the protection of moral and material interests resulting from any scientific, literary or artistic production of which he is the author	right to strike	right to control one's health and body
right to freedom indispensable for scientific research and creative activity		right to be free from interference
right and duty to develop its culture		right to a system of health protection
		right to an equal opportuniy to enjoy the highest attainable level of health
		right to work
		right to social security
		right to sical insurance
		right to be free from hunger

Table 4: revised taxonomy of human rights (lower nodes)

However, as some of these terms are sequences of words (in some cases more than 10), they are not likely to produce matches when running searches on existing corpora. Therefore, it would be useful to have a wider range of search patterns for

each term. Let us consider some of the civic rights presented in Table 3, for example two civic rights and their correspondent violations:

term	alternative terms	alternative search pattern
right to life	right of life	rigth_life
homicide	homicides, murder, murders, crime, crimes	
freedom of speech	freedom of expression, lack of censorship, free speech	freedom_speech lack_censorship free_speech
censorship	censorships, censoring, bowdlerization, media blackout, news blackout	media_blackout news_blackout

Table 5: Example of expansion of the lower nodes of the revised taxonomy

Each of the terms in column one may be expressed in different ways in the targeted dataset. Column two provides alternative terms, including the plural equivalent, as well as some possible synonyms or hyponyms. Finally these terms may be expressed in sentences in which there is a gap between these terms, and therefore contain an underscore character.

Named Entity Recognition with Spacy

A possible method for detecting these terms in a text or set of texts is Named Entity Recognition (NER). The NLP library Spacy, which includes a pre-trained model with a number of entity types such as Person, Product or Date, also allows to train the recognizer with customized examples, or to create a new model from the beginning. The Spacy website provides step by step instructions on how to train the recognizer to locate and label, as well as re-usable code.

Although It would take time to train to locate all the terms in our reviewed taxonomy, is still possible to start the task and train some of the examples appearing in Table 5. Table 6 below provides four terms and the entity labels that we want them to carry, which are abbreviations of the subclasses that contain these terms:

Subclass	Entity Label	term
Civic right	CR	right to life
Violation of civic right	VCR	homicide
Civic right	CR	freedom of speech
Violation of civic right	VCR	censorship

Table 6: Examples of four terms and their labels for the Spacy NER recognizer

Figure 1 shows a screenshot of the method to update the NER recognizer in Spacy,

```
TRAIN_DATA = [
    ("Who is Shaka Khan?", {"entities": [(7, 17, "PERSON")]}),
    ("I like London and Berlin.", {"entities": [(7, 13, "LOC"), (18, 24, "LOC")]}),
]
```

Figure1: screenshot of the training data to update the NER recognizer in Spacy

On the basis of the method in Figure 1, we have trained the recognizer to detect the terms 'human rights', 'right to life', 'freedom of speech and censorship'. Except for the term 'human rights', which we label as HR, the entity labels for the other 4 terms are provided in Table 5 above. The terms and their corresponding labels are highlighted in the code below, (the .py file will be also attached with this email):

TRAIN_DATA = [
    ("respect **human rights**.", {"entities": [(8, 20, **"HR"**)]}),
    ("Article 2 of the human rights Act protects your **right to life**.", {"entities": [(48, 61, "CR**"**)]}),
    ("**freedom of speech** is a principle that supports the freedom of an individual or a community to articulate their opinions and ideas without fear of retaliation, censorship, or legal sanction.", {"entities": [(0, 17, **"CR"**)]}),
    ("**censorship** is the suppression of speech, public communication, or other information.", {"entities": [(0, 10, **"VCR"**)]}),
    ("the Global Study on **homicide** has been expanded into a special six-booklet format.", {"entities": [(20, 28, **"VCR"**)]}),
]

As shown in figure two below, the words have been detected and labelled as entities correctly:

```
Entities [('human rights', 'HR')]
Tokens [('respect', '', 2), ('human', 'HR', 3), ('rights', 'HR', 1), ('.', '', 2)]
Entities [('right to life', 'CR')]
Tokens [('Article', '', 2), ('2', '', 2), ('of', '', 2), ('the', '', 2), ('human', '', 2),
('rights', '', 2), ('Act', '', 2), ('protects', '', 2), ('your', '', 2), ('right', 'CR', 3)
('to', 'CR', 1), ('life', 'CR', 1), ('.', '', 2)]
Entities [('homicide', 'VCR')]
Tokens [('the', '', 2), ('Global', '', 2), ('Study', '', 2), ('on', '', 2), ('homicide',
'VCR', 3), ('has', '', 2), ('been', '', 2), ('expanded', '', 2), ('into', '', 2), ('a', '',
2), ('special', '', 2), ('six', '', 2), ('-', '', 2), ('booklet', '', 2), ('format', '', 2)
('.', '', 2)]
Entities [('freedom of speech', 'CR')]
Tokens [('freedom', 'CR', 3), ('of', 'CR', 1), ('speech', 'CR', 1), ('is', '', 2), ('a', ''
2), ('principle', '', 2), ('that', '', 2), ('supports', '', 2), ('the', '', 2), ('freedom',
'', 2), ('of', '', 2), ('an', '', 2), ('individual', '', 2), ('or', '', 2), ('a', '', 2),
('community', '', 2), ('to', '', 2), ('articulate', '', 2), ('their', '', 2), ('opinions',
'', 2), ('and', '', 2), ('ideas', '', 2), ('without', '', 2), ('fear', '', 2), ('of', '',
2), ('retaliation', '', 2), (',', '', 2), ('censorship', '', 2), (',', '', 2), ('or', '',
2), ('legal', '', 2), ('sanction', '', 2), ('.', '', 2)]
Entities [('censorship', 'VCR')]
Tokens [('censorship', 'VCR', 3), ('is', '', 2), ('the', '', 2), ('suppression', '', 2),
('of', '', 2), ('speech', '', 2), (',', '', 2), ('public', '', 2), ('communication', '', 2)
(',', '', 2), ('or', '', 2), ('other', '', 2), ('information', '', 2), ('.', '', 2)]
```

Figure 2: screenshot of the Python console after executing the code

Concluding remarks

After providing a temporary solution for the exercise, we can conclude the creation and testing of a taxonomy or ontology is most likely a protracted task. Although we may have provided some improvements on a possible taxonomy for the semantic fields of human rights, it is only by testing this taxonomy that we can see these improvements, and perhaps further improve it. Next steps would include a wider and deeper training of the recognizer as well as some tests on a dataset.