

Documentation for Prerequisite Relation Annotation

TelDH Lab

1 Content of the Document

The present document contains the instructions and recommendations for performing manual annotation of prerequisite relations in educational materials.

The novel annotation protocol presented within this document is the result of a virtuous cycle of modeling, testing and revision of annotation guidelines and specifications that reached its third version, which is here presented.

Intended Audience

This document is intended especially for researchers working in the field of AI and Education that want to study how prerequisite relations are established between concepts in educational materials. However, the annotation protocol could be used also by teachers that want to create (collaborative) educational concept maps representing the propaedeutic relations that are established between the concepts mentioned in a textbook.

Terms and Definitions

Here you can find the definition of the terms we will use along the document.

- **Prerequisite Relations:** also called PRs, they are pedagogical relations established between educational concepts described in educational materials. These relations express what should be understood first in order to avoid knowledge gaps when learning a new topic.
- **Concept:** a concept is an abstract and general idea conceived in the mind. In education, concepts represent the building blocks of learning, namely what a student should understand in order to acquire new knowledge. Depending on learner's needs with respect to granularity, concepts can be very general (e.g., algebra, geometry, mathematics etc.) or very specific (e.g., radius, integer multiplication, fraction denominator). Either way, they are represented in texts as lexical entities constituted by a single or multi-word term.

- **Corpus:** a corpus is a document, the textual resource, such as a textbook or any learning material in written form. It can be enhanced with labels to become an annotated text, i.e. a text where certain information related to its content is made explicit through annotation.
- **Annotation:** in general terms, annotation consists of the process of adding comments, notes, explanations, or other types of external marks that can be attached to a (part of a) document. In this document, we refer to annotation as the manual process led by humans consisting of adding labels to a *textual corpus* in order to indicate the presence of prerequisite relations between two *concepts* mentioned in the corpus.
- **Annotation Protocol:** guidelines and specifications aimed at indicating how to obtain corpora enriched with explicit information regarding a certain phenomenon and that are designed to be reproduced on any unannotated texts at any time.
- **Annotation Project:** the set of tasks aimed at building an annotated text that includes explicit annotations about the phenomenon being studied (here, prerequisite relations in educational texts). Within the project it is included and defined every element involved in the annotation, such as the documentation concerning annotation specifications, the corpus to be annotated, the people involved in the process and the results obtained.
- **Project Manager:** the person or team leading the annotation project. The manager is in charge of taking decisions concerning the goals and settings of the annotation project.
- **Annotation Guidelines:** instructions and recommendations that indicate how to perform annotation. They can be consulted at any time during annotation, but should be known in advance.
- **Annotator:** the person that performs the annotation on the corpus according to the guidelines and project principles.
- **Gold Standard Dataset:** the output of the annotation project. A Gold Standard Dataset (or *Gold-PR dataset* with reference to a Gold Standard annotated with PR relations) is a ground truth annotated dataset based on a single trusted annotation, or obtained by combining multiple manual annotations into a single one.
The Gold-PR are generally exploited to:
 - i) obtain informative analysis of the annotated phenomenon;
 - ii) test the performances of machine learning systems trained using it;
 - iii) compare the gold labels against those obtained using automatic systems to test their accuracies.

2 Prerequisite Annotation Protocol

The annotation protocol is aimed at building resources manually annotated with prerequisite relations. To this aim, the protocol is designed to support manual identification of the relations while reading educational materials, for example a textbook. The ultimate goal of this process is the creation of Gold-PR datasets, i.e. manually labelled sets of items resulting from the annotation of a single expert or from the combination of all annotators' judgements. These resources are highly valuable for training and testing automatic systems and to investigate the main aspects involved in the realisation of PR relations in textual data.

Our approach, as we will discuss shortly, is specifically designed to annotate PRs on educational texts aiming to identify manually recognised relations between concepts and, at the same time, being able to retrieve PR linguistic realisations in texts. The latter goal is particularly relevant for prerequisite relation identification, if we consider that the content of a textbook is designed to guide students through a pre-arranged learning path designed to tackle relevant concepts and highlight their relations. However, multiple sequences can be legitimately proposed for the same path (as clear from the multiple textbooks concerning the same topic). As an example, consider when an author decides to explain a topic starting from broad concepts and definitions as opposite to cases when (s)he starts from specific cases or examples: top-down and bottom-up approaches have a significant impact on the representation of prerequisites between concepts since they imply opposite relations (i.e. from general to specific or from specific to general).

Having an annotation strictly bound to the text has the advantage of making the annotation independent from any external resource and, at the same time, it reflects the teaching approach of the author of the resource being annotated. As a consequence, the novel PR annotation approach defined by our protocol allows the creation of resources that can be used to investigate concepts organisation within the content of a learning material and if the PRs appear within recurrent linguistic patterns.

We will now present the main steps of an annotation project that follows the principles of our annotation protocol.

2.1 Before Annotation - Preliminary Decisions

Prior to the text annotation phase, a *project manager*, who leads the annotation project, should define the setting of the annotation tasks that will be carried out. Based on the annotation protocol, the project manager shall be responsible for taking decisions about the following aspects.

- **Annotation Purpose:** define the reason for starting the annotation project, which consists of deciding how the annotations will be used once they are created by annotators;
- **Textual corpus:** choose the resource to be annotated and prepare it for annotation (e.g., perform OCR if your text comes in PDF format, clean

the text from undesired information, check if any pre-processing step is needed);

- **Annotation setup:** decide where the annotation should be carried out, for example one might want to use an annotation tool like PRET¹. In that case, the manager should take care of any possible preliminary setup required by the chosen tool;
- **Annotators:** decide who should perform the annotation task. For instance, one might want to restrict the annotators recruitment to only those having certain characteristics, such as a certain degree or age. Then, recruit the annotators, explain them the task and set up a pilot study to assess their understanding of the protocol principles;
- **Revision:** decide whether and how to revise the annotations after completing the annotation task;
- **Gold Dataset Creation:** whether and how to combine multiple annotations in order to obtain the Gold-PR dataset to be used as Ground Truth data.

Note that, although Revision and Gold dataset creation are addressed after annotation, the manager should take decision about them in advance.

Once these elements are defined, the next step is performing annotation.

2.2 PR Annotation

The annotation protocol for annotating PRs between concepts in educational materials specifies the following steps for the annotation process:

1. Find the relevant domain concepts from the educational resource;
2. Read the text and, if you encounter a concept that needs some prior knowledge to be understood, indicate its prerequisite concept(s) from the list of concepts found at step 1.
3. Revise the pairs you created reading again the portion of text where they were annotated.

Next to these quite simple principles, we provide annotation guidelines (see 3.1) to tackle problems that could be encountered during annotation. We will now describe each step more in detail.

¹<http://telldh.dibris.unige.it/pret/>

Find concepts. The first step of the annotation protocol consists of identifying domain concepts in the text. This step is tackled as an autonomous step of the annotation process and the project manager is in charge of taking preliminary decisions concerning the way concepts should be extracted and used. We recommend to adopt one of the following approaches:

- a) Let annotators identify domain concepts while annotating the text with PR relations;
- b) Provide a pre-selected list of validated concepts that the annotators have to use as-it-is;
- c) Provide a pre-selected list of validated concepts that the annotators can refine and update during the annotation.

Consider that, with option a), letting annotators freely add concepts without exchanging views with the other annotators produces richer but less homogeneous annotations. If the manager decides to adopt option b) or c), (s)he should take care of extracting the concepts from the text. We do not provide in the protocol recommendations about which strategy should be preferred for concept extraction: this choice is left to the manager.

Annotate PRs. Our annotation task consists of pairing the concepts mentioned in the text if the annotator considers them as having a prerequisite relation while reading an educational text, i.e. following the flow of the text to identify concepts that are used by the author to explain a new concept. We ask experts not to label the absence of a PR, since it would make the annotation process hard to carry on (one would have to label at least $n(n-1)$ PRs, with n being the number of concepts).

Revise Annotations. Manual annotation is known to contain errors due to misinterpretation of the text or the guidelines, but also to natural distraction. To address this problem and improve the internal coherence of the annotations, the protocol includes an ‘in-context revision’, namely a revision phase to be carried out after the annotation.

Revision allows to check if the pairs created by an annotator should be kept in the annotation, modified or deleted. In order to revised his/her pairs, each annotator is required to read again the portion of text where she/he found the PR relation to check if a pair was inserted intentionally or by mistake. In case reducing the workload of this step is a priority, we recommend to revise at least those PRs identified by a low number of annotators.

2.3 After Annotation - Evaluation and Combination

Once annotation and revision are completed, the project manager should take care of the subsequent steps concerning the annotations reliability evaluation and combination.

Agreement Evaluation Agreement between two or more annotations is a measure of similarity and annotation reliability. Cohen’s and Fleiss’ kappa are two of the most prominent metrics used for agreement evaluation and we recommend to use them on PR annotated texts in the following cases: the first is used between two annotations selected by you in the analysis interface, the latter is computed between all annotations produced for the same text. The agreement value ranges from 0 to 1, where 1 identifies perfect agreement. Although one can apply the metrics as preferred, we propose an alternative computation strategy that we specifically designed to fit the peculiarities of PR annotations. Specifically, we propose to apply the k metric to PR-annotations produced according to our protocol taking into account the transitive property of PRs.

More in detail, we assume that two annotators agree on the PR $A \prec C$ in both the following cases: *i)* both annotators manually created the pair $A \prec C$ and *ii)* one annotator created the pair $A \prec C$ and the other created the pairs $A \prec B$ and $B \prec C$.

The metric is thus computed as follows. Given the list T of concepts used during annotation, consider as total items of the annotation task the list P of each pairs-wise combination p of concepts in T , regardless the relation direction (i.e., $A \prec B$ and $B \prec A$ are both included in P). For each annotator, consider as positive PR each p that is either manually created by the annotator or that can be derived for the transitive property. Consider p as a negative PR otherwise. Then, compute k for each pair of annotators using the following equation, where P_o stands for the observed agreement (i.e., probability for an item to receive the same annotation by both raters), while P_e denotes the agreement expected by chance (i.e., the probability of each individual category)..

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \quad (1)$$

Gold Standard Dataset Creation Gold standard datasets are intended to provide a generally accepted annotation of a phenomenon that can be looked at as accurate and reliable reference. Therefore, gold datasets are usually built by a pool of experts by performing a shared annotation or, more frequently, by combining their single annotations. Different criteria can be adopted to combine the annotations, ranging from intersecting to joining them. The former approach means taking only those annotations inserted by all the annotators, while the latter consists of including the annotations added by each annotator. The first approach maximises the precision with respect to the ground truth, while the latter maximises the recall. The former can be useful when the annotators are not expert, while the latter is suggested when the annotators are experts and the boundaries of the phenomenon have some fuzziness that are to be taken into account. In between of these two approaches, others can be defined that weight the number of annotators or specific features.

3 Annotation Manual

The annotation specifications comprise the annotation instructions and recommendations for performing PR annotation on texts, systematised within the *annotation manual*. The annotation manual is composed of two complementary resources: the *Annotation Guidelines* (AG), whose aim is to describe how the annotation process should be carried out in order to reduce inconsistencies in the annotations, and a list of *Knowledge Elicitation Questions* (KDE), aimed at clarifying dubious cases through direct questions and examples. Guidelines concern different issues of the annotation process:

- i) Concept identification issues (addressed by AG rec. 1-3);
- ii) Text annotation issues (AG rec. 3-6);
- iii) PR features and properties (AG rec. 7-9);
- iv) Annotation revision (AG rec. 10-12).

We will now discuss each of those issues and how they are addressed in the protocol. KDE are aimed at clarifying dubious cases through direct questions and examples and at helping the annotators think over hard cases.

3.1 Annotation Guidelines

1. The goal of the annotation is identifying a prerequisite relation between two distinct terms of a textual corpus. The two terms represent concepts described in the text and can be referred to as target and prerequisite concepts.
2. A concept can be either a single or multi-word term extracted from the corpus.
3. Insert a prerequisite relation for a target concept if you think you need to know the information related to a different concept in order to understand what you are reading about the target concept. Each of the two concepts must be present either in the initial Terminology provided by the project manager or in the manual terminology built by you (i.e. the annotator) during the annotation process. If a concept is still missing in the terminology, add the corresponding term and then insert the relation (if this is allowed by the project manager).
4. The relation must be inserted exactly in the context (i.e. the sentence) where you find it. A concept could be mentioned more than once along the text, each time introducing novel information and recalling different concept(s). Make sure to add the prerequisite relation between two concepts exactly where the target concept description recalls the knowledge related to the concept you identified as prerequisite.

5. Build a concept pair only if a prerequisite relation does exist between the two: if you think that a relation between two concepts does not occur in the text, do not insert any relation.
6. *Trust the text*: you must annotate only concepts and relations that can be acquired from the text. Do not consider concepts and relations recalled from your background knowledge about the topic.
7. A concept cannot be a prerequisite of itself: self prerequisites such as "computer is a prerequisite of computer" will not be allowed by the system.
8. Do not introduce loops in the annotation. Imagine that you have already annotated that: i) "fruit" is a prerequisite of "citrus", and ii) "citrus" is a prerequisite of "orange". By annotating that "orange" is a prerequisite of "fruit", you will create a loop.
9. Every time you insert a relation you must also define its weight. Allowed values comprise: *strong* (the prerequisite is absolutely necessary to understand the other term) and *weak* (the prerequisite is very useful but not strictly necessary).
10. After completing your annotation, you should also perform an annotation check aimed at revising you previously created PRs. By reading again the portion of text where you entered a relation, decide whether you want to confirm, delete or modify the pair.
11. Delete a prerequisite relation if you added it by mistake. Keep in mind, however, that you can delete one single instance of a pair at a time: if the same pair is annotated with the prerequisite relation in another part of the text, that will be preserved. If you think that ANY prerequisite relation between two concepts should be deleted, you must delete each of the relations having those two concepts.
12. Modify a prerequisite relation if you assigned it the wrong weight. You can modify the weight of the relation if you believe that text expressed a different relation strength than the one you originally assigned to the PR when creating it.

3.2 Knowledge Elicitation Questions

1. Which concepts (among those mentioned in the text) you need to master in order to understand the meaning of the target concept?
2. Which concepts are recalled to define the target concept?
3. Are other concepts mentioned in the same context (e.g., sentence or paragraph) of the target concept? If so, are they useful to understand the meaning of the target concept?

4. Does the target concept represent a special case of another concept mentioned in the text (e.g., *circumference*[target] is a special case of *ellipse*[prerequisite])?
5. Does the target concept show a part-of relation with another concept mentioned in the text (e.g., the *elbow*[target] is a part of an *arm*[prerequisite])?
6. Does the target concept consists of sub-elements already mentioned in the text (e.g., *elbow*, *forearm* and *shoulder*[prerequisites] are parts of the *arm*[target])?
7. Is the target concept caused by another previously described concept (e.g., *rain*[prerequisite] causes *floods*[target]) or vice versa (e.g., *rain*[target] is caused by *low pressure*[prerequisite])? If so, which one? Try to follow the relation proposed by the text author to understand if a prerequisite relation exists.