PRAT Quickstart Guide

Index

Introduction

PRAT is the first online tool designed to support prerequisite relation annotation on textbooks.

Thanks to PRAT you can:

- Upload your text and annotate prerequisite relations between the concepts described in it to create an annotated dataset;
- Try different automatic prerequisite extraction methods already implemented in PRAT;
- Compare the results of automatic prerequisite extraction methods with your manually created dataset;
- Analyse and visualize the annotated datasets.

In this quickstart guide we'll take you through the PRAT functionalities and show you how you can build your first annotated dataset and perform your analyses.

Start PRAT tool

Login

Start PRAT Tool and log in using your credentials. If you are not registered yet, create a new account by clicking on *"Not registered yet?"* (see figures below).
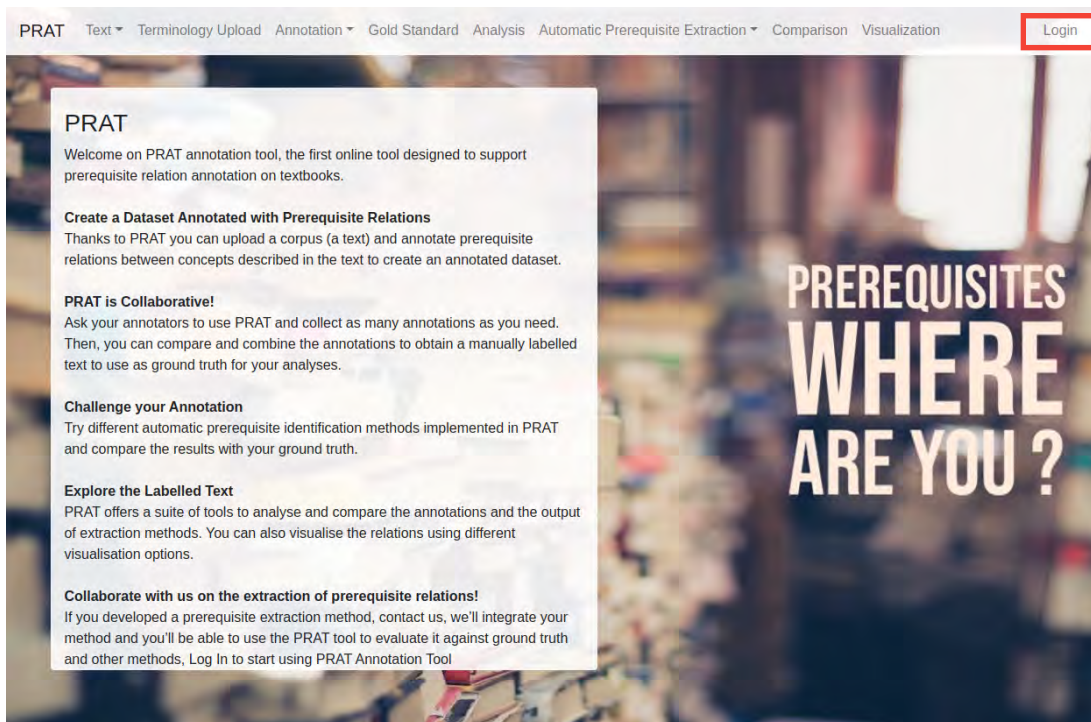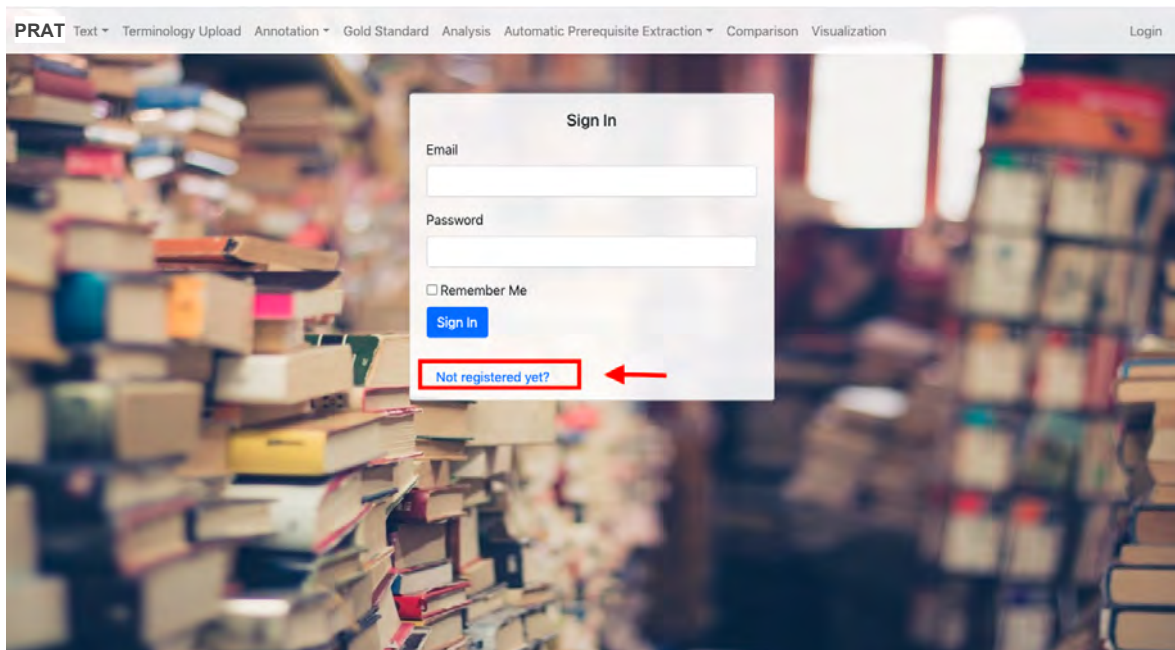

Figure 1: PRAT home and login button

Figure 2: PRAT new account registration

Text Upload

Upload a text you want to annotate to PRAT and provide the required metadata (text must be in .txt format). "Browse" and select the corpus file from your pc: the system will load the text and you will see it as a preview on the interface (see Figure 3).

Figure 3: PRAT text upload module

Add Text Structure

Many texts have a hierarchical structure represented by titles, sub-titles and paragraphs. In order to add the structure to your text, select section titles using checkboxes. Once you finish, save the file.
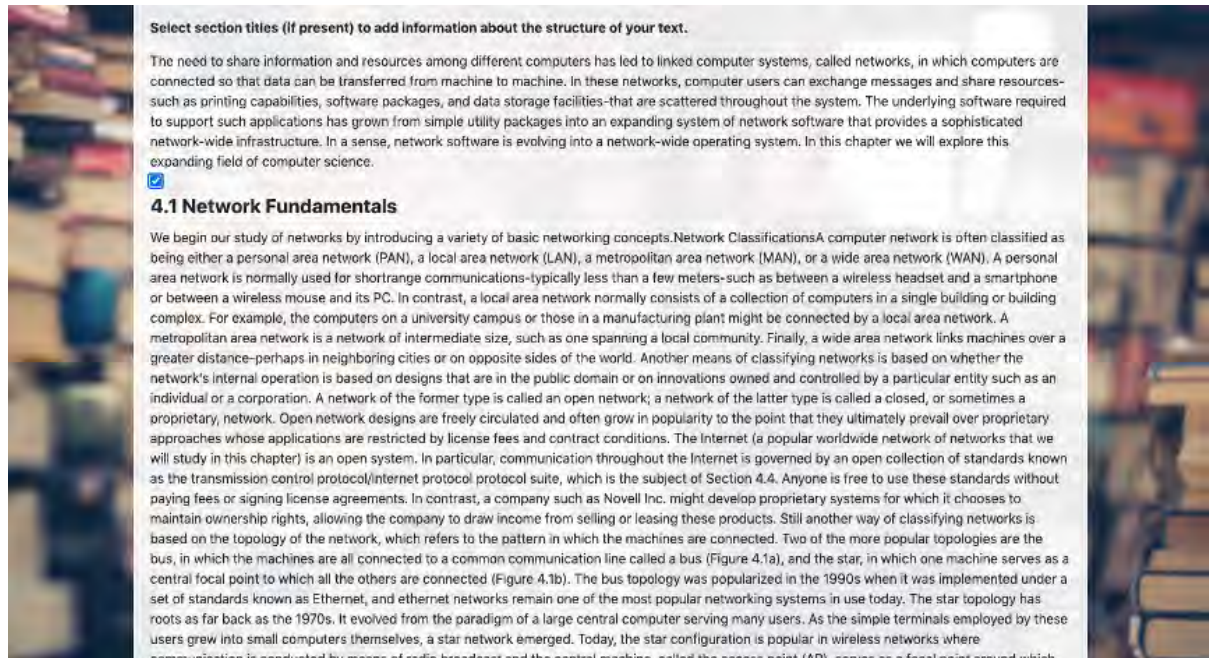

Figure 4: Paragraph selection

Download and Delete

Within the Text Module you can also delete the text and download the linguistic analysis in CoNLL-U Format automatically performed on your text using UDPipe[1].

Terminology Upload

Upload a file containing the most relevant terms mentioned in your text. The file should contain one term on each line and, for each term, its lemma (e.g., *'dog'* for *'dogs'*, *'to be'* for *'is'*).

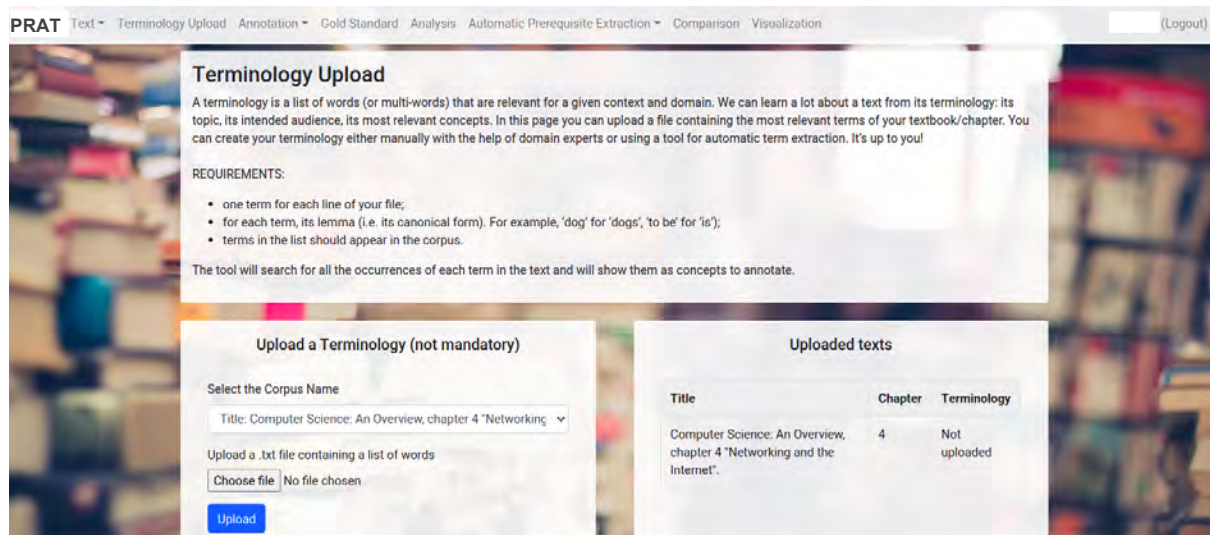[1] Pre-trained using the English model version 2.5.

Figure 5: Terminology upload module

Dataset Annotation

Once loaded the .txt file and its associated terminology, within the Annotation Module you can start the annotation of prerequisite relations on your text.
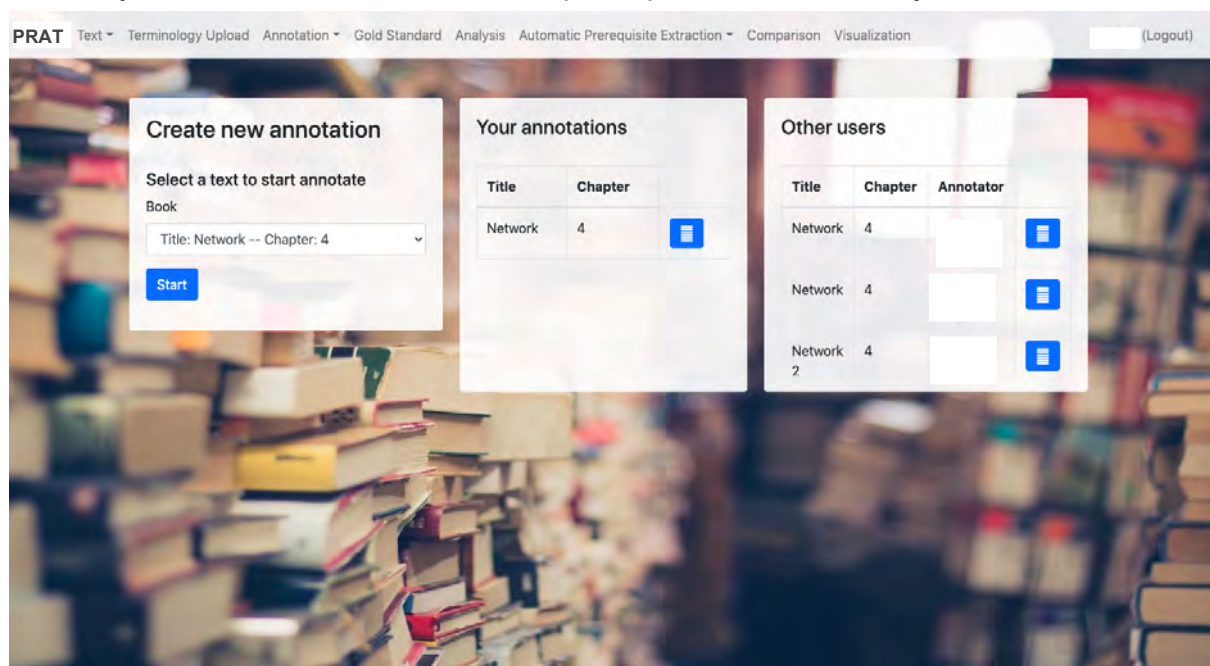

Figure 6: PRAT Annotation module

Annotation Guidelines

Read the **Annotation Guidelines** before starting the annotation to know how to spot when two concepts are connected by a prerequisite relation.

Create Relations between Concepts

Consider Figure 7 below. By double-clicking on the "computer" concept within the *Text* area, a window will open where you can type the term that you identify as its prerequisite concept (e.g., "network"). You should also define the *weight* of the relation. Again, consult the Guidelines to know which weight should be assigned to each relation.
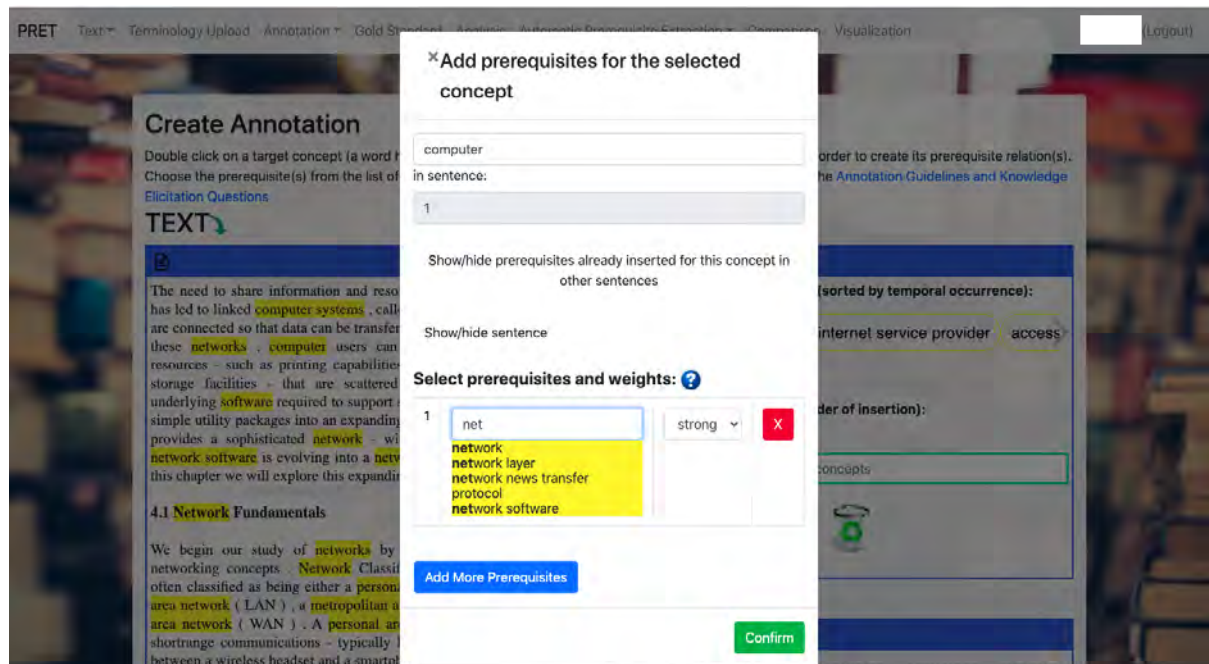

Figure 7: Prerequisite relation dialog box

Add New Concepts

In order to manually add new concepts not included in the uploaded terminology, highlight the word - for example "machine" - and drag it to the green box on the right.
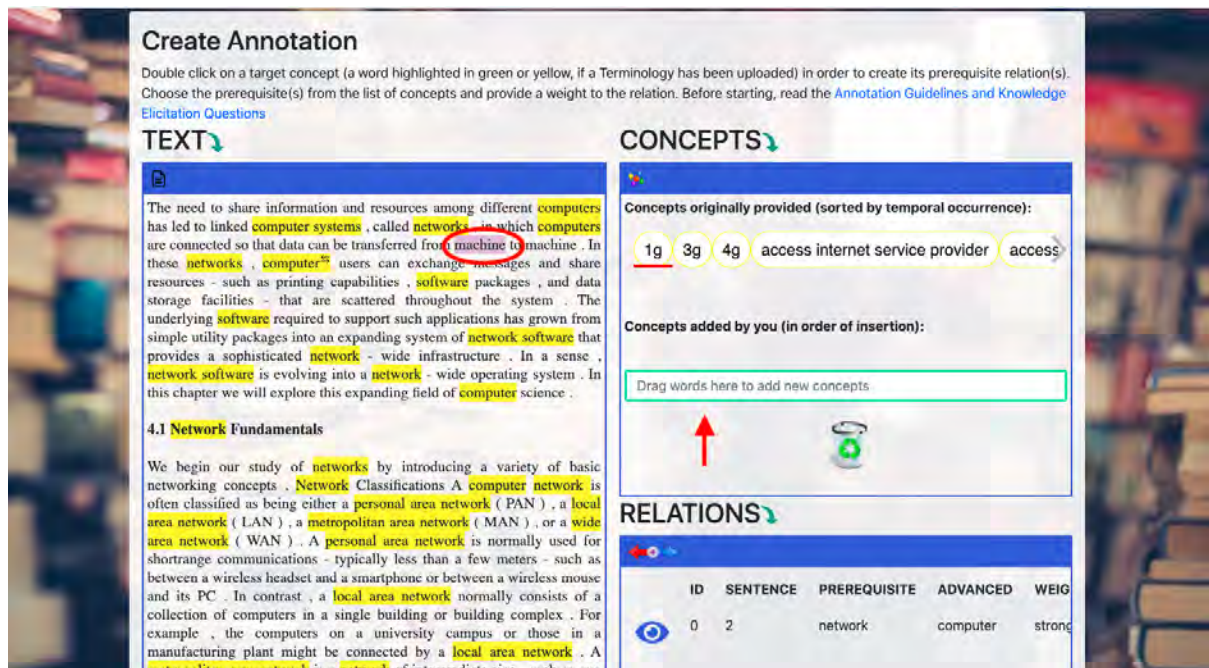
Figure 8: new concepts insertion box

Delete Concepts

If you change your mind or realise you have made a mistake adding a new concept, you can delete it by clicking on the *Bin* icon (see the image below) and typing the concept name into the dialog box.
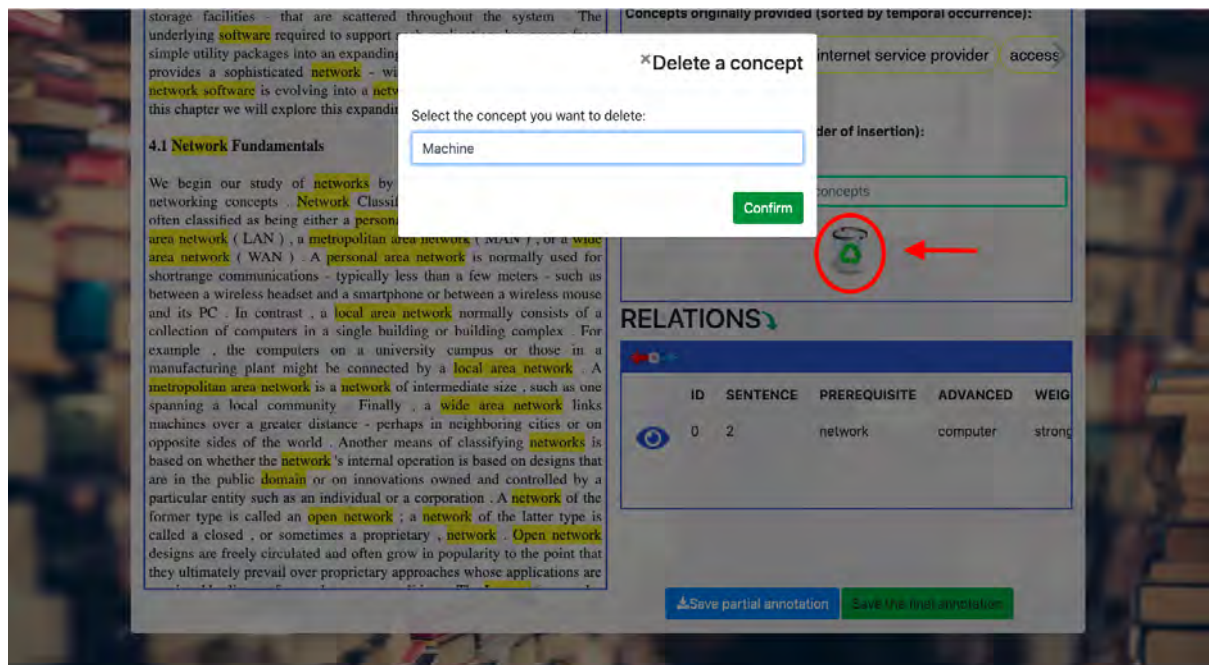

Figure 9: new concepts deletion

Save Annotation

By clicking on the "*Save Annotation*" button, a window will appear where you can choose whether to save a draft of your work (that you can resume later) or permanently save your final annotation. Note that by choosing the latter option you won't be able to add more relations, but you'll still be able to modify those already inserted (see next section for more details about the revision process).
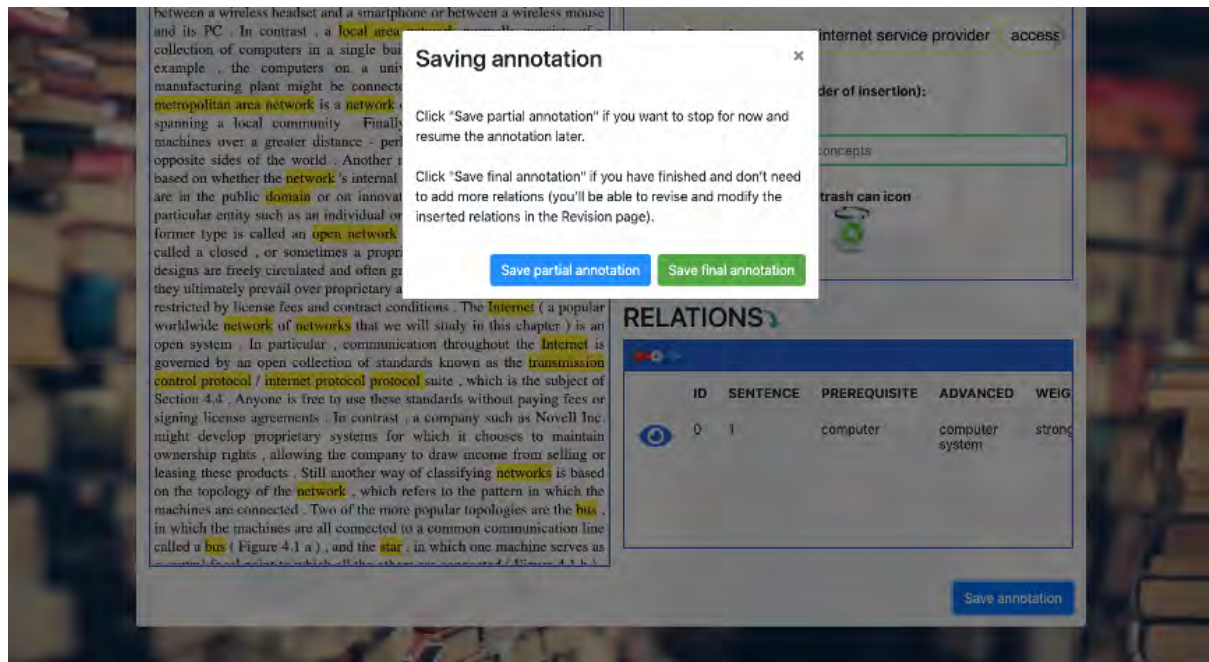

Figure 10: saving annotation dialog box

Annotation Revision

An annotation can be revised after its final version has been saved on PRAT. The **Revision** section within the **Annotation** module allows you to delete or modify the pairs in your annotations. Revision is useful both to refine your manual annotation and to identify in which cases you inserted an annotation error.

Figure 11 shows the revision interface. For each pair in your annotation, you can choose one of these three options: Confirm (in green), Delete (in red) or Change Weight (in yellow). To help you choose the most suitable option, the system shows you the portion of text where you inserted the relation. If you decide to confirm (or delete) a relation, you can also indicate the relation type (or the motivation for deleting it) by choosing among the provided labels (column *Tag*). Read the instructions before you begin to know the meaning of each label. The change of weight will happen automatically.

If you want to revise exclusively those pairs identified by you only (thus not present in the annotation of other possible annotators), check "*Show only unique relations*".

Figure 11: PRAT annotation revision

Gold Standard Dataset Creation

Gold Standards are ground truth annotated datasets that can result from a single annotation considered as gold or as from the combination of multiple annotations on the same text.

Consider Figure 12. To create a gold standard, select a text from the 'Book' menu, pick out the annotations you want to use for your gold standard dataset, a combination criterion, set a name for your gold dataset (default option is combining annotators' names) and then press the **Create Gold Standard** button. Gold datasets will be visible and available for download in the *"List of Golds"* box.

If you want to know more information and suggestions regarding the **Combination Criteria**, consult the info box at the top of the PRAT tool page.

Figure 12: gold standard generation

Annotation Analysis

The analysis module provides tools to analyze annotations - either gold or from a single annotator - or the output of automatic methods for prerequisite extraction (see next section for details about the methods). First choose a text from the *"Book"* menu then select which type of analysis you want to see.

Figure 13: Analysis module interface

Data Summary

Data summary shows you quantitative analysis of the selected annotation, either considering prerequisite pairs and the graph of relations that can be obtained from that annotation.

Linguistic Analysis

Linguistic analysis allows you to search for pairs in the annotation that match specific criteria that you set in the interface (e.g., show all relations where "computer" is prerequisite and the weight is *weak*) and see them in their context.
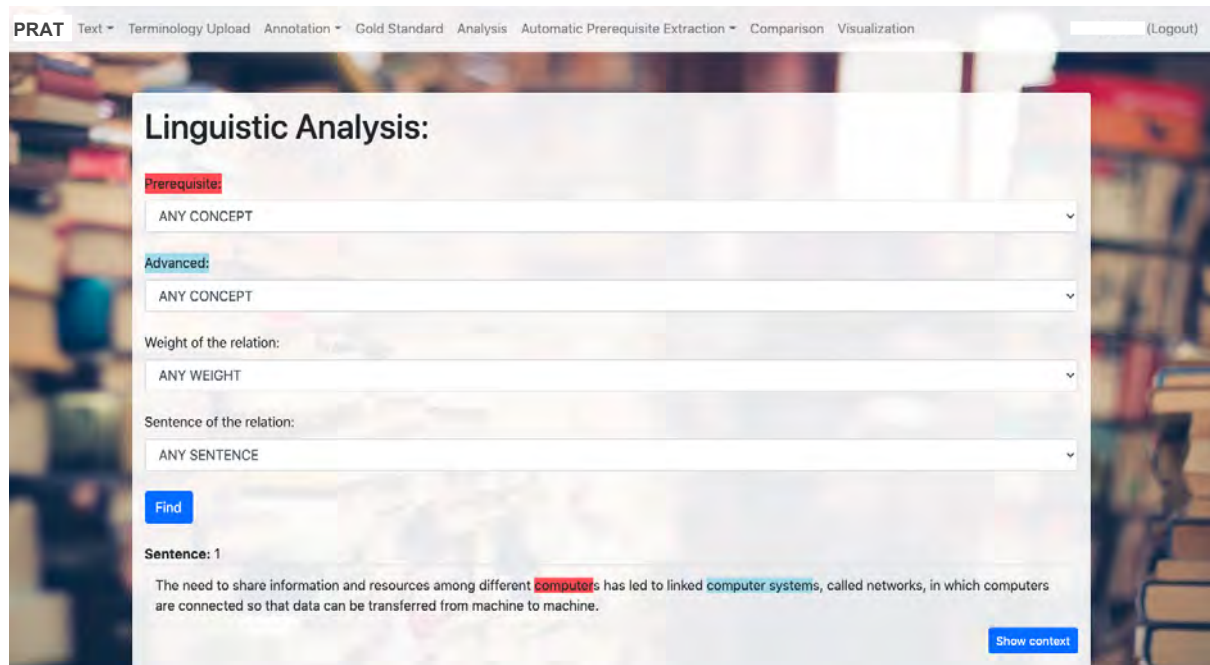

Figure 14: an example of linguistic analysis

For each relation you can also explore their context (*next and previous sentence*), the linguistic analysis (dependency syntactic tree) of the sentence where the relation was entered (*POS*) and the relation graph of the two concepts involved in the relation (*Graph*).
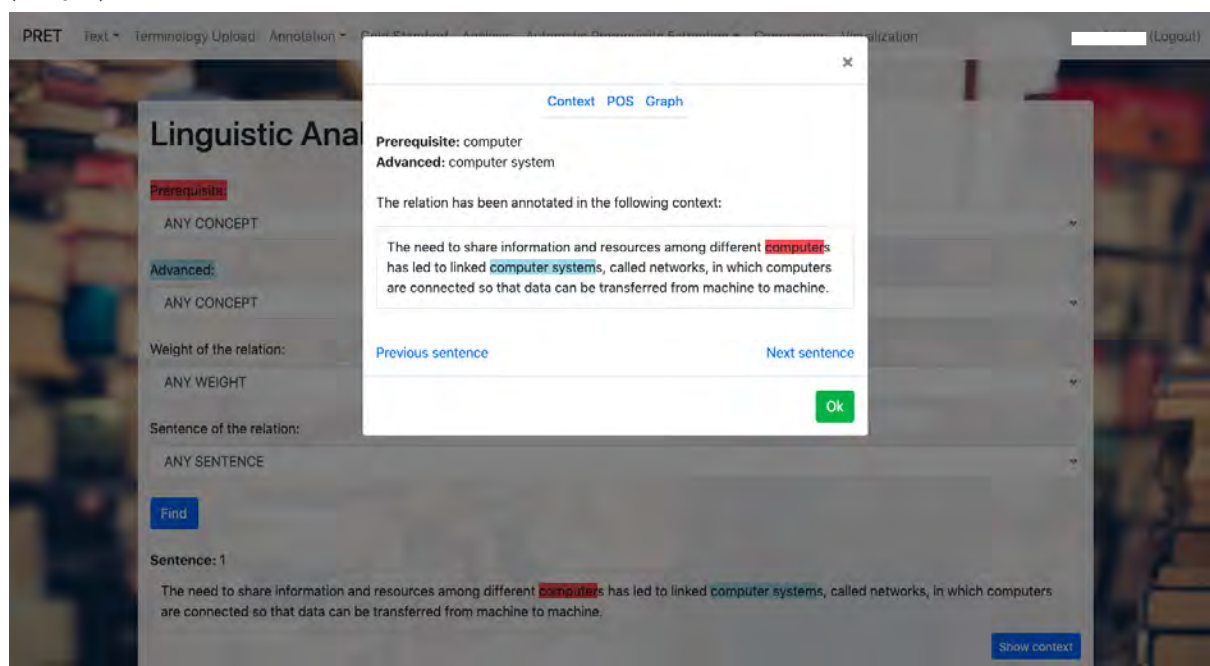

Figure 15: context dialog box

Agreement Calculation

Agreement between two or more annotations is a measure of similarity and annotation reliability. PRAT computes Cohen's and Fleiss' kappa: 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.

Prerequisite Extraction Methods

This module offers various methods for the automatic extraction of prerequisite relations. Read the **Methods Description** before you begin in order to understand how each automatic extraction method works.

As you can see from Figure 16, select the text from the "*Book*" menu, choose a method and launch. For method 3.Relational Metric, 5.Textbook Structure and 6.Temporal Patterns, you need to define a value that will be used as a threshold to discriminate between prerequisite and non-prerequisite pairs.
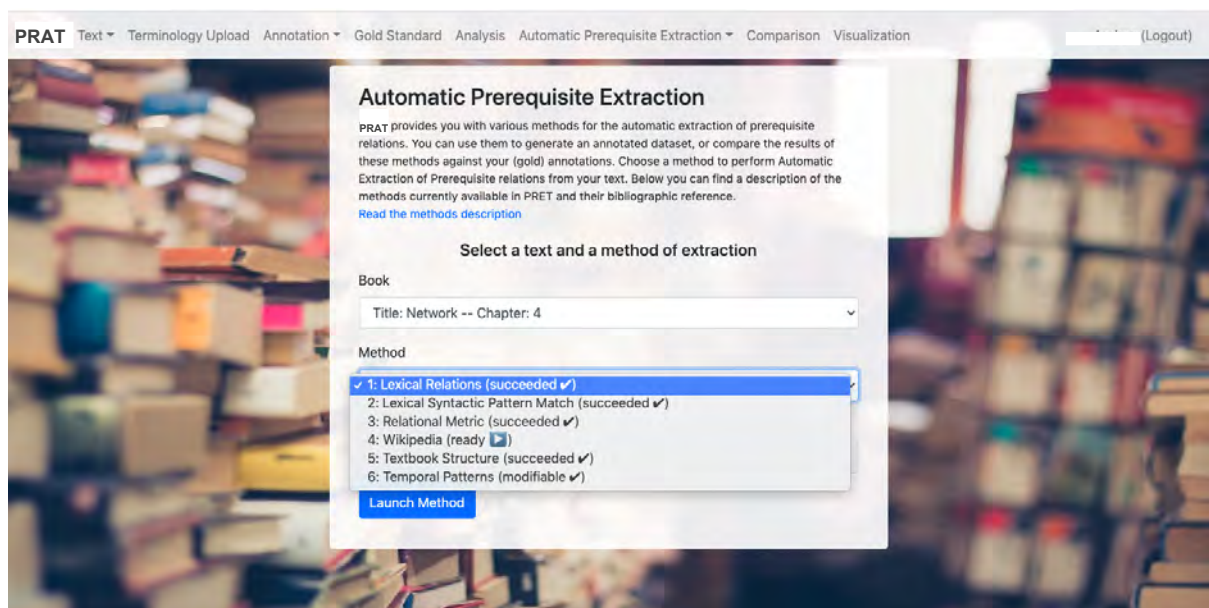
Figure 16: Automatic Prerequisite Extraction module

Figure 17: an example of threshold insertion

View Results

To consult and download the results you need to access the **View results** section (see an example in Fig.17). Using the *"Filter"* button you can search for relations involving specific concepts, for example all relations having "computer" as prerequisite (Figure 19), or all relations having "spyware" as target (Figure 20).

Figure 18: an example of results and download

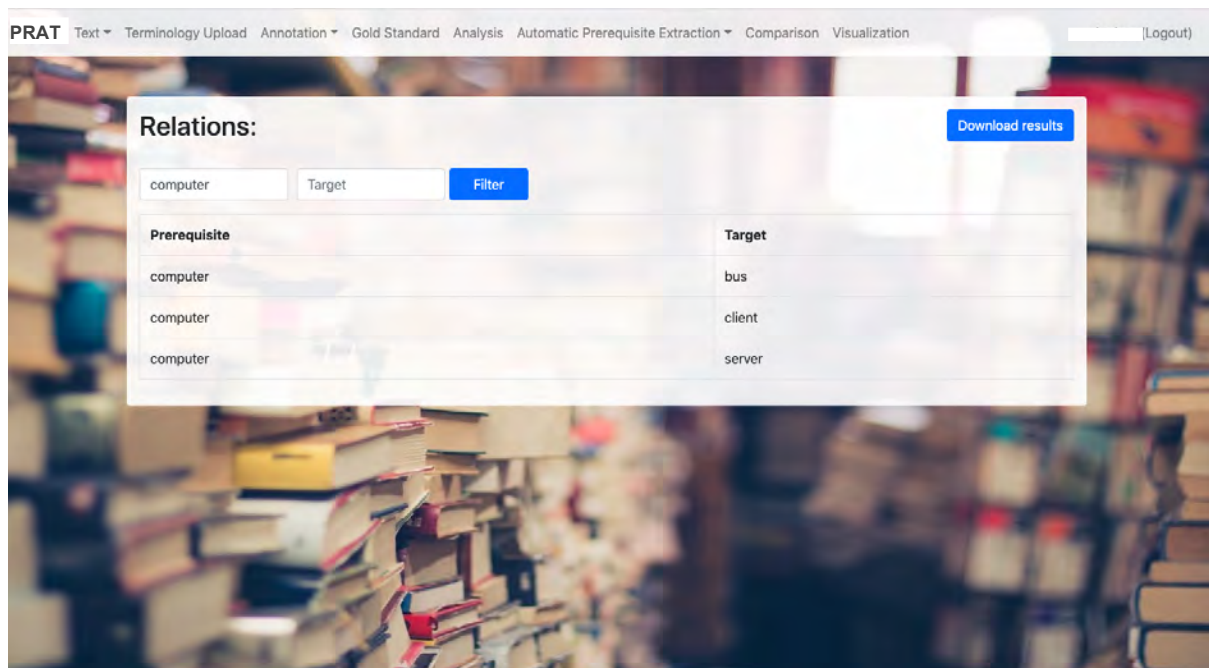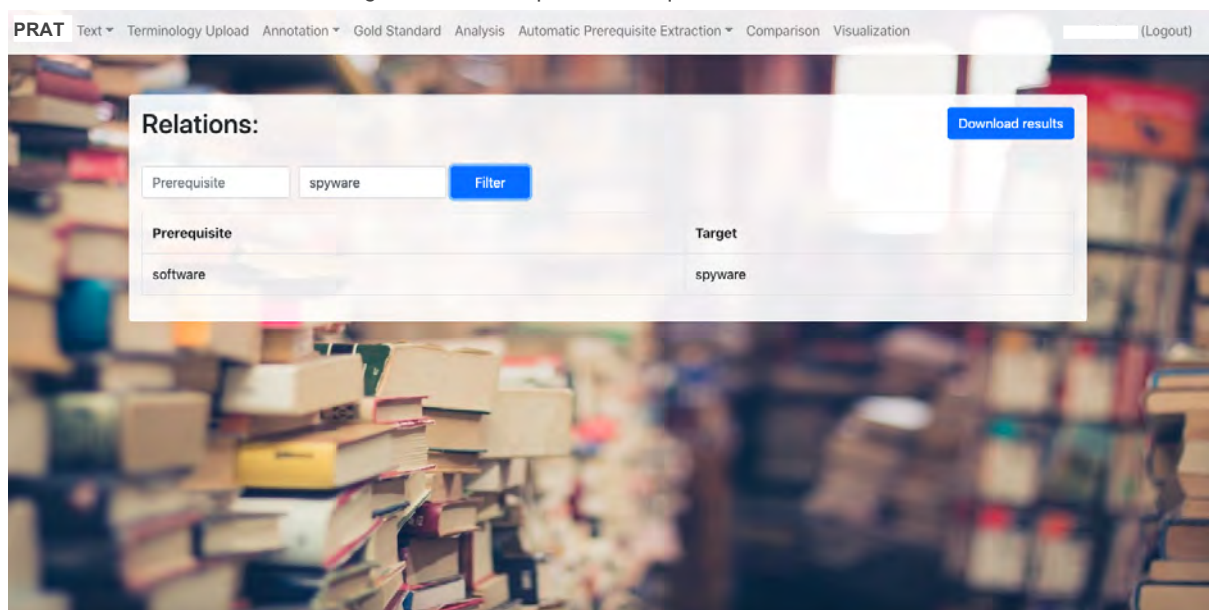Figure 19: an example of "Prerequisite" filter results

Figure 20: an example of "Target" filter results

Annotation Comparison

This module allows to compare two datasets to see their similarities and differences. Select the text in the "*Book*" menu, then as Item 1 and 2 you can choose a single annotation, a gold standard or a result of an automatic extraction method. Then Launch Comparison!
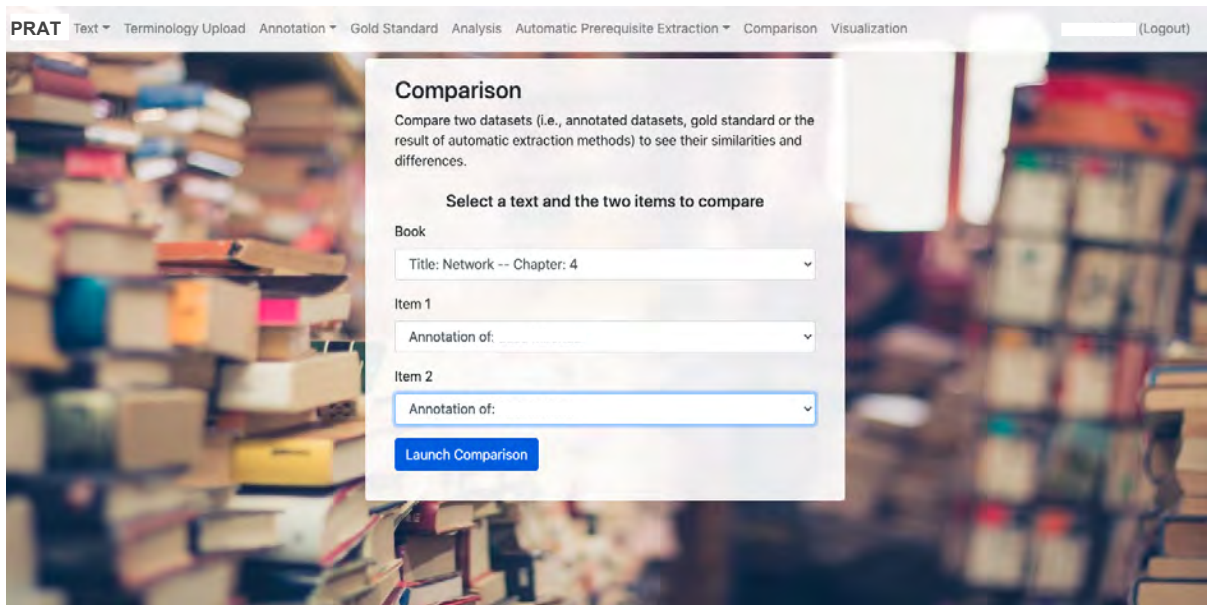
Figure 21: Comparison module

Summary

Summary tab provides a comparison of the characteristics reported in Data Summary (Annotation Analysis module) for each of the annotations under comparison.

Shared Relations
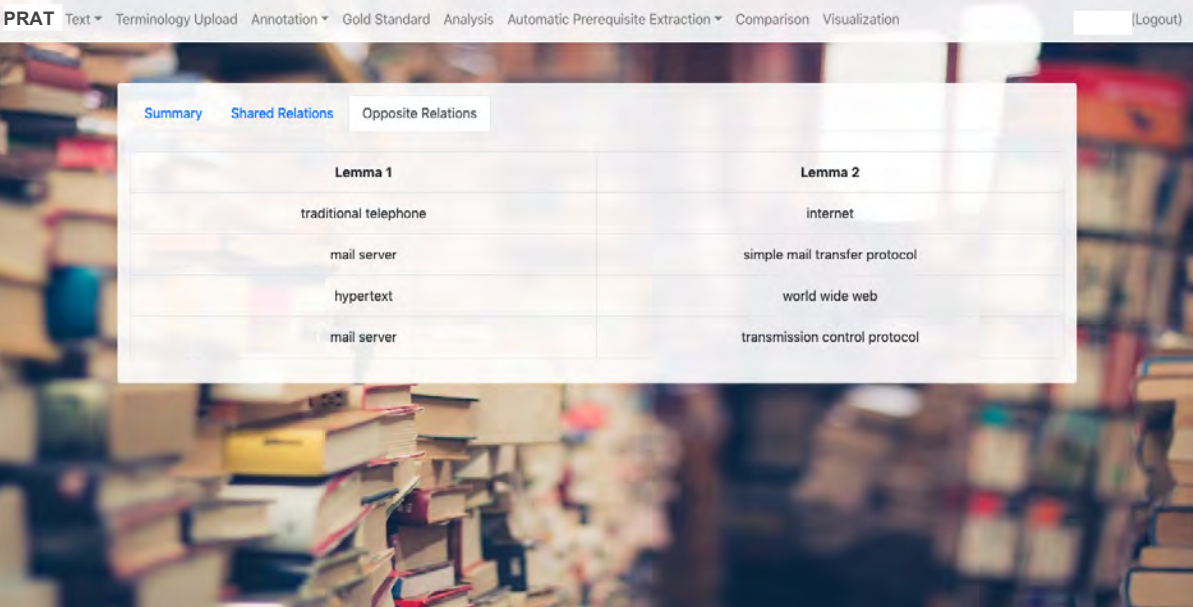
Shared Relations tab (Figure 22) shows which pairs appear in both annotations, regardless of their weight which is shown aside.

Prerequisite	Target	Weight Luca	Weight Alessio
computer	network	strong	strong
software	network software	strong	strong
network	personal area network	strong	strong
network	local area network	strong	strong
network	metropolitan area network	strong	strong
network	wide area network	strong	strong
network	internet	strong	strong
network	bus	strong	weak
network	star	strong	weak
bus	ethernet network	strong	strong

Figure 22: Shared Relations

Opposite Relations

Opposite Relations tab (Figure 23) shows which relations appear in both annotations, but with inverted roles for concepts. See the image below: in the relation between *traditional telephone* and *internet*, for one of the two annotators *traditional telephone* is prerequisite of *internet*, for the other the opposite is true.

Figure 23: Opposite relations example

Annotation Visualisation

Within this module you can generate customized visualisations, according to your needs. Select the text in the "Book" menu, the annotation to visualise and the visualisation choosing among those available. Only for the Temporal Pattern method you will be able to choose the Gantt Graph.
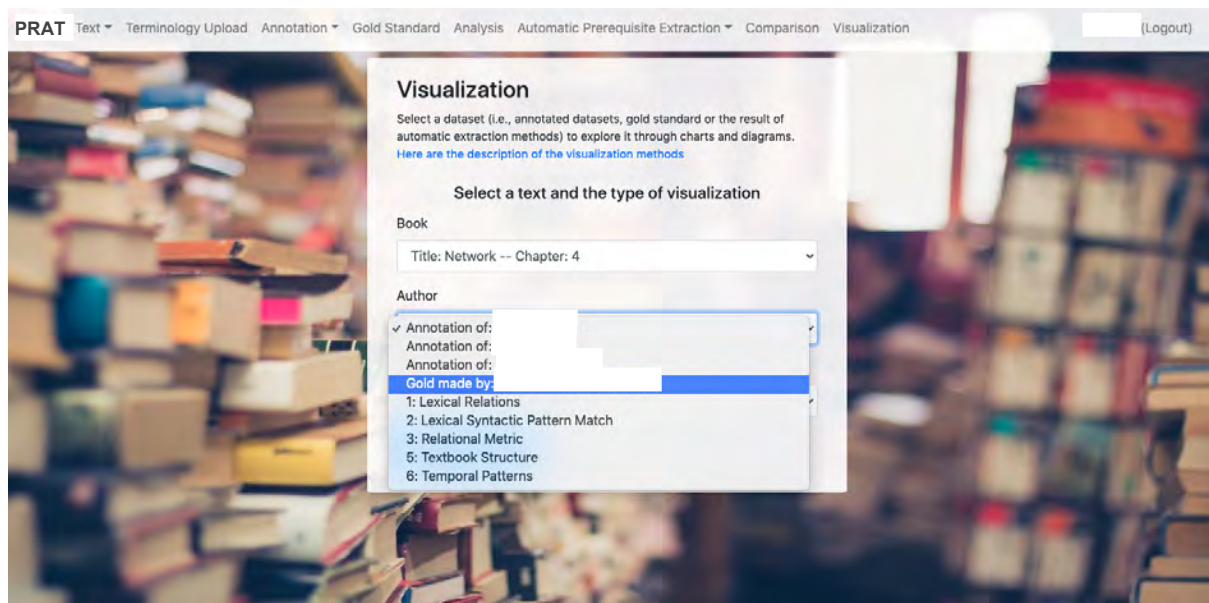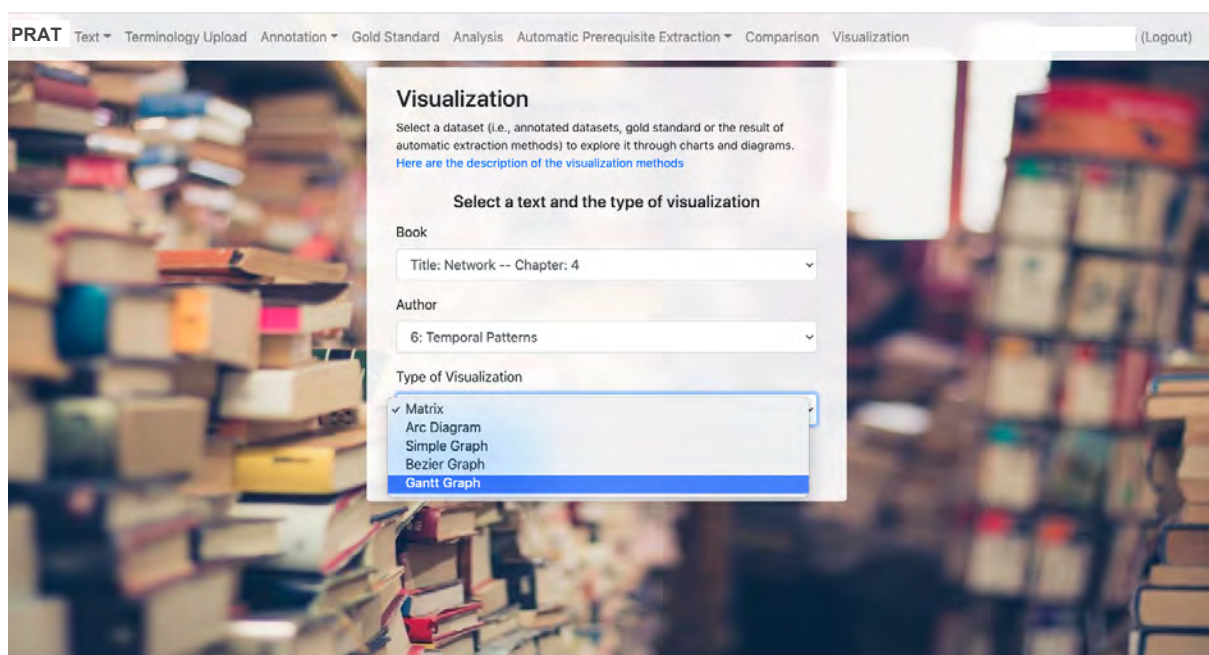
Figure 24: Visualisation module

Figure 25: Gantt Graph for Temporal Patterns method

Matrix

In the Matrix visualisation concepts of the terminology are displayed along the horizontal and vertical axes and their intersection is coloured if a prerequisite relation exists between the two concepts. You can choose to distinguish the colours based on the relation weight as it was assigned during annotation: red if weight is strong, green if it's weak. Concepts on the axes can also be ordered according to different criteria: alphabetical order (*name*, default option), *frequency* of occurrence

in the text, aggregating concepts involved in a similar number of relations (*cluster*) and *temporal appearance* along the text progression.
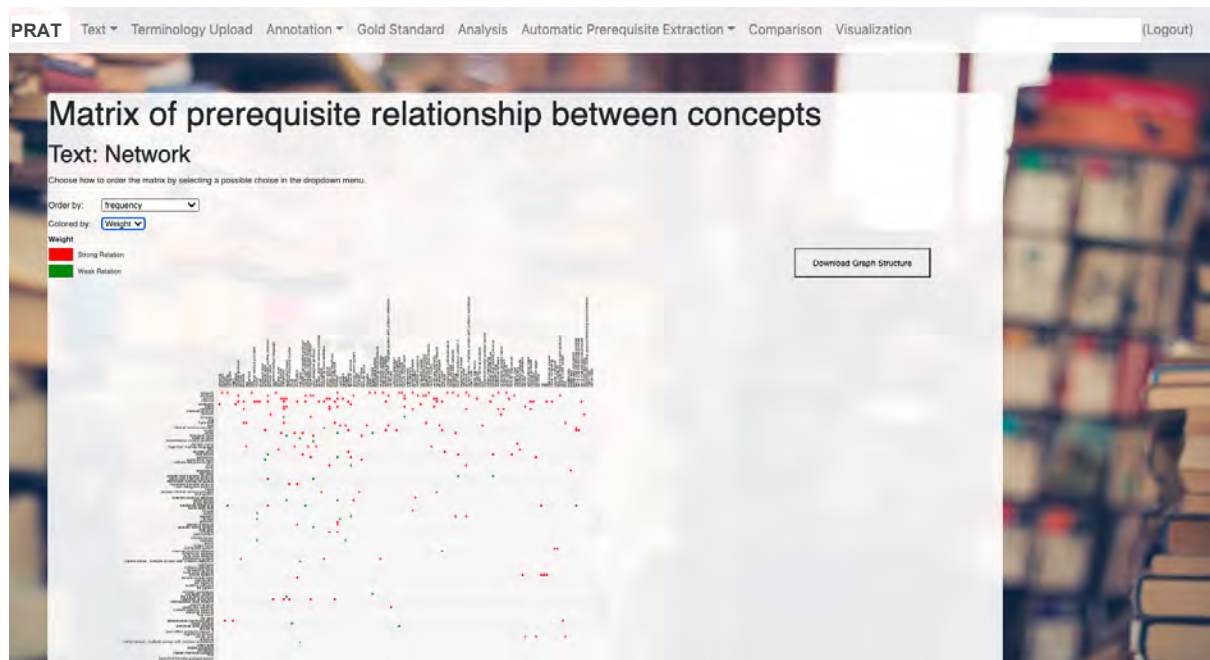

Figure 26: Matrix Visualisation

Arc Diagram

The Arc Diagram visualisation is a flat representation of a Matrix: concepts are displayed along the horizontal axis and arcs represent prerequisite relations connecting them. You can filter which pairs you want to display by choosing the section of the book where they were inserted and the relations direction (i.e., forward if the prerequisite concept is mentioned in the text before the target concept, backward otherwise).
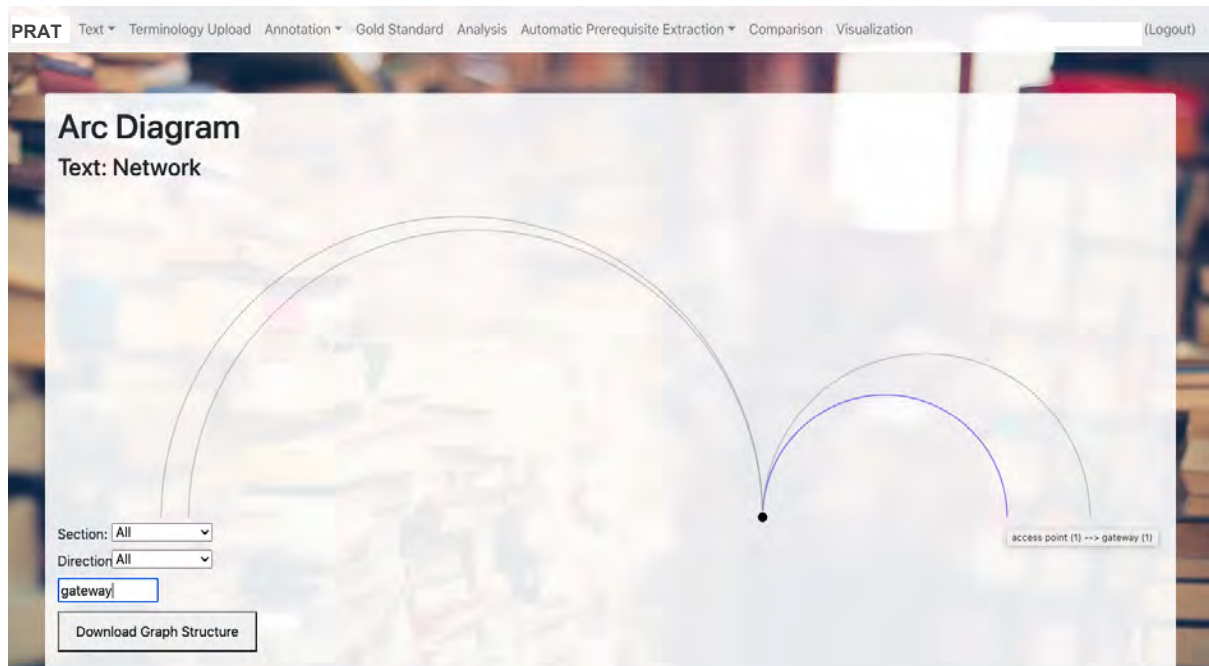
Figure 27: Arc Diagram Visualisation

Graphs

Graphs represent concepts as nodes and prerequisite relations as edges. You can represent the graph in two different styles (there is no difference in the graph structure): simple (as in the image below) or Bezier.
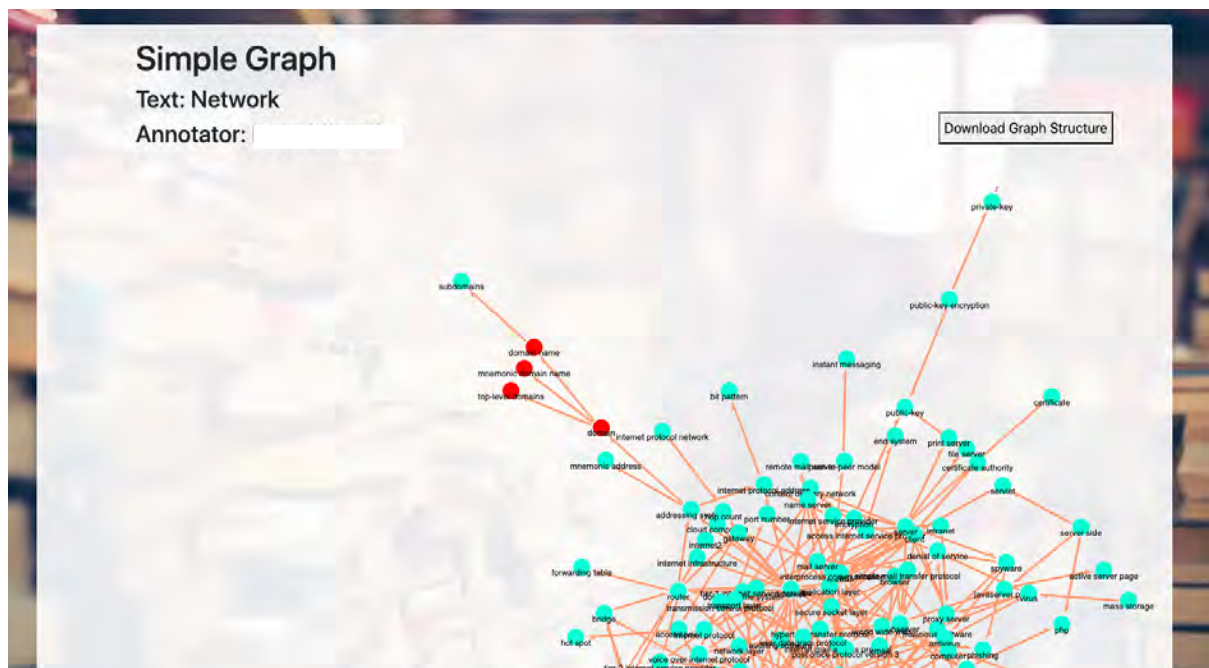

Figure 28: Graph Visualisation

Gantt Graph

Gantt diagram shows intervals of text where concepts are relevant along the horizontal temporal axis (time is measured as a progression of sentences), while concepts are arranged along the vertical axis, according to their temporal appearance in the text. Yellow blocks are used when a concept has a unique relevance interval; green and red blocks refer to the first and last interval of relevance respectively; blue blocks are those between a green and red block. Blocks become purple when selected.
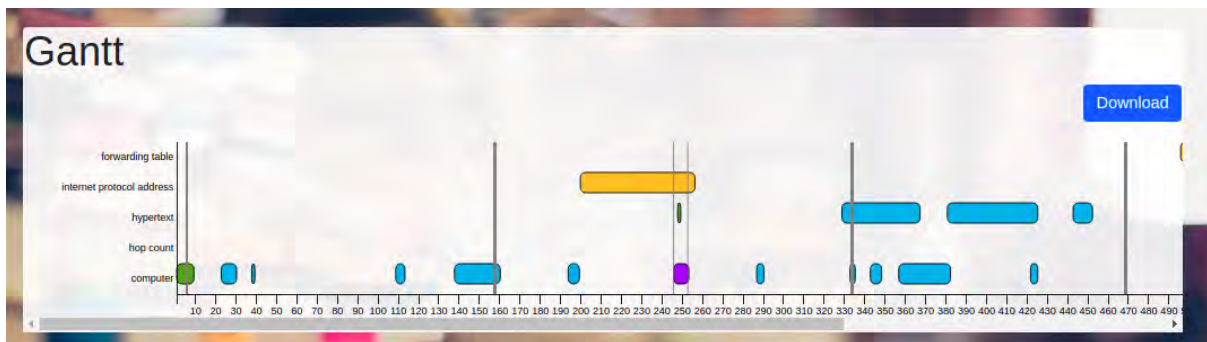

Figure 29: Gantt Graph Visualisation