Section RDFization

What is annotated and how

- (similar to biotéa) but using SPAR: similar to http://www.essepuntato.it/2014/doco/example
    - The sections themselves.
        - Section type using deo and sro ontologies (e.g. type deo:Introduction)
        - TODO:
            - create a "section-type" property. Currently the section type is linked its parent-structure simply like this: [parent] rdf:type [sectionType].
            - Create a section-type class for sections that could not be classified.
    - The order of the textual structural elements using the Co-owl-ontology is rdfized. Description see *Ciccarese: "The Collections Ontology: creating and handling collections in OWL 2 DL frame"* (http://www.semantic-web-journal.net/system/files/swj432.pdf). Here you also find SPARQL-query examples on how to utilize inferencing.
    - The paragraphs
        - (same as in biotea): type doco:Paragraph, c4o:hasContent [text].
    - Citations. I.e. sentences/fragments that cite a reference.
        - in-text frequency (e.g. how often is reference r38 cited?)
        - location in text: (e.g. part of paragraph [p]. [p] part of [section]. [section] type doco:Introduction.)
        - Why is a citation in the running text annotated?
            - We can classify the citation.
            - We can rank citations according to frequency (in one article, in whole dataset (TODO find out: possible to do aggregate SPARQL 1.1 queries to add in-text citation counts across several articles?)
            - An application can highlight the portion of text that forms the citation for the user
            - If in addition, this portion of text or a superordinate structure (the paragraph, the section) contains annotations from controlled vocabularies or keywords, maybe this citation can be assoicated with the concept/keyword and thus reveal more about the content of the cited resource or at reveal information about the relation between the citing document and the cited resource.

Extensions to the SPAR-ontologies

- They contained only an intransitive has-part relationship (the same holds for the inverse object-property: „part-of").
    - Created an owl ontology that extends the doco and pattern ontologies by adding a superordinate transitive has-part and part-of object property (see src/main/resources/zpid_doco.owl).

Queries

Query for the sections directly contained by the article

```
prefix c4o: <http://purl.org/spar/c4o/>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>
```

```
prefix co: <http://purl.org/co/>
prefix deo: <http://purl.org/spar/deo/>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix po: <http://www.essepuntato.it/2008/12/pattern#>
prefix doco: <http://purl.org/spar/doco/>
prefix fabio: http://purl.org/spar/fabio/

#select the first-level sections of the article and the text contained in these sections.
select  ?sec ?p ?text
where {
  ?article rdf:type fabio:Article.
  ?article prism:doi "10.5964/ejcop.v3i1.23".
  ?article po:contains ?list.
  ?list co:element ?sec.
  ?sec rdf:type doco:Section.
  ?sec po:contains ?listSubSecAndParagraphs.
  ?listSubSecAndParagraphs co:element ?p.
  ?p rdf:type doco:Paragraph.
  ?p c4o:hasContent ?text.
}
```

Query for all sections (directly contained sections and transitively contained subsections) of the article

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix doco: <http://purl.org/spar/doco/>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>
prefix co: <http://purl.org/co/>
prefix zpid_doco: <http://www.zpid.de/zpid_doco#>
prefix po: <http://www.essepuntato.it/2008/12/pattern#>

select  ?article  ?sec
where {
  ?article prism:doi "10.5964/ejcop.v3i1.23".
  ?article zpid_doco:containsIndirectly ?list.
  ?list rdf:type po:Structured.
  ?list co:element ?sec.
  ?sec rdf:type doco:Section.
}
```

Query that ranks the citations of references of an article according to frequency

```
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>
prefix cito: <http://purl.org/spar/cito/>
prefix biro: <http://purl.org/spar/biro/>
prefix c4o: <http://purl.org/spar/c4o/>

#The top 10 references cited in the article with in-text citation frequency > 1.
Select ?ref ?inTextCount
where {
```

```
  ?art prism:doi "10.5964/ejcop.v3i1.23".
  ?art cito:cites ?refText.
  ?ref biro:references ?refText.
  ?ref c4o:hasInTextCitationFrequency ?inTextCount
  FILTER(?inTextCount > 1)
} ORDER BY DESC (?inTextCount ) LIMIT 10
```

Query to retrieve the titles of the top 10 references cited in the article with in-text citation frequency > 1.

```
prefix cito: <http://purl.org/spar/cito/>
prefix dcterms: <http://purl.org/dc/terms/>
prefix biro: <http://purl.org/spar/biro/>
prefix c4o: <http://purl.org/spar/c4o/>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>

Select ?refTitle ?inTextCount
where {
  ?art prism:doi "10.5964/ejcop.v3i1.23".
  ?art cito:cites ?refText.
  ?ref biro:references ?refText.
  ?ref c4o:hasInTextCitationFrequency ?inTextCount.
  ?refText dcterms:title ?refTitle.
  FILTER(?inTextCount > 1)
} ORDER BY DESC (?inTextCount ) LIMIT 10
```

Query to retrieve the Top cited authors of an article.

```
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix doco: <http://purl.org/spar/doco/>
prefix cito: <http://purl.org/spar/cito/>
prefix dcterms: <http://purl.org/dc/terms/>
prefix biro: <http://purl.org/spar/biro/>
prefix c4o: <http://purl.org/spar/c4o/>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>
prefix pro: <http://purl.org/spar/pro/>


#The top cited authors of the article.
Select (sum(?inTextCount) AS ?count) ?authorName ?firstName
where {
  ?art prism:doi "10.5964/ejop.v8i3.308".
  ?art cito:cites ?reference.
  ?inTextCitation biro:references ?reference.
  ?inTextCitation c4o:hasInTextCitationFrequency ?inTextCount.
  ?reference dcterms:title ?refTitle.
  ?role pro:relatesToDocument ?reference.
  ?role pro:withRole pro:author.
  ?group pro:holdsRoleInTime ?role.
  ?group foaf:member ?member.
  ?member foaf:familyName ?authorName.
  ?member foaf:givenName ?firstName.


} GROUP BY ?authorName ?firstName ORDER BY Desc (?count)
```

Query to retrieve the sections ranked by number of citations occurring within their paragraphs.

```
prefix doco: <http://purl.org/spar/doco/>
prefix cito: <http://purl.org/spar/cito/>
prefix dcterms: <http://purl.org/dc/terms/>
```

```
prefix biro: <http://purl.org/spar/biro/>
prefix c4o: <http://purl.org/spar/c4o/>
prefix prism: <http://prismstandard.org/namespaces/basic/2.0/>

prefix po: <http://www.essepuntato.it/2008/12/pattern#>


#The sections with the most citations within an article.
Select (count(?ref) as ?refcount) ?sec
where {
  ?art prism:doi "10.5964/ejcop.v3i1.23".
  ?art cito:cites ?refText.
  ?ref biro:references ?refText.
  ?refPtr c4o:denotes ?ref.
  ?refPtr c4o:hasContext ?para.
  ?sec po:contains ?para.

} GROUP BY ?sec ORDER BY desc (?refcount)
```