

Text Graph Ontology

A Semantic Web approach to represent
genetic scholarly editions



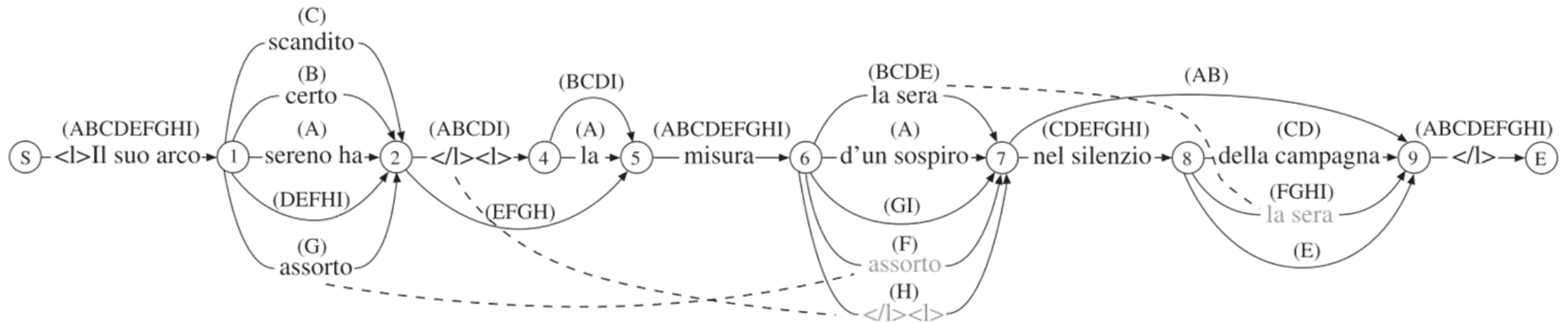
<https://github.com/Wolkenstein/tgo-ontology>

Text Graph Models

Graphs? What graphs?

- Several projects and models used graphs to represent textual variation in the recent years
 - GODDAG: Sperberg-McQueen and Huitfeldt 2004
 - GrAF: Ide and Suderman 2007
 - Schmidt and Colomb 2009
 - CollateX (Haentjens Dekker et al. 2015)
 - “The Stemmaweb Project” 2012–
 - Efer 2016
 - TAG: Haentjens Dekker and Birnbaum 2017

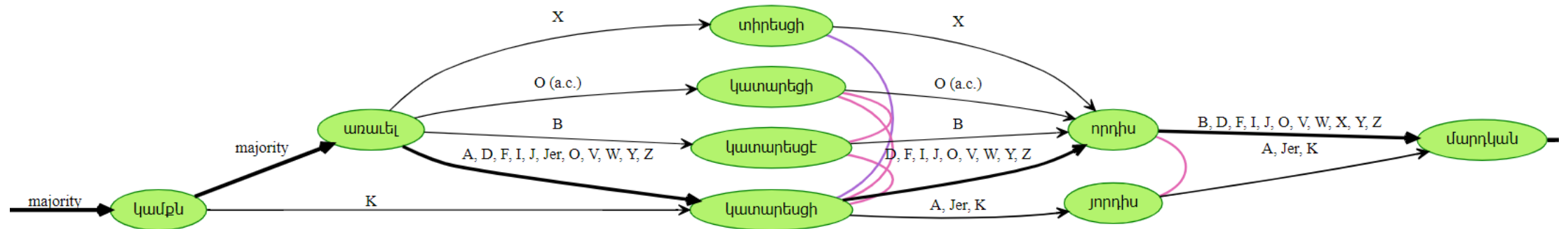
Schmidt, Desmond and Robert Colomb. 2009.
 “A Data Structure for Representing Multi-Version Texts Online.”
International Journal of Human-Computer Studies 67 (6): 497–514



Genesis of a sentence of Valerio Magrelli's poem *Campagna Romana*. [Schmidt/Colomb 2009: 502]

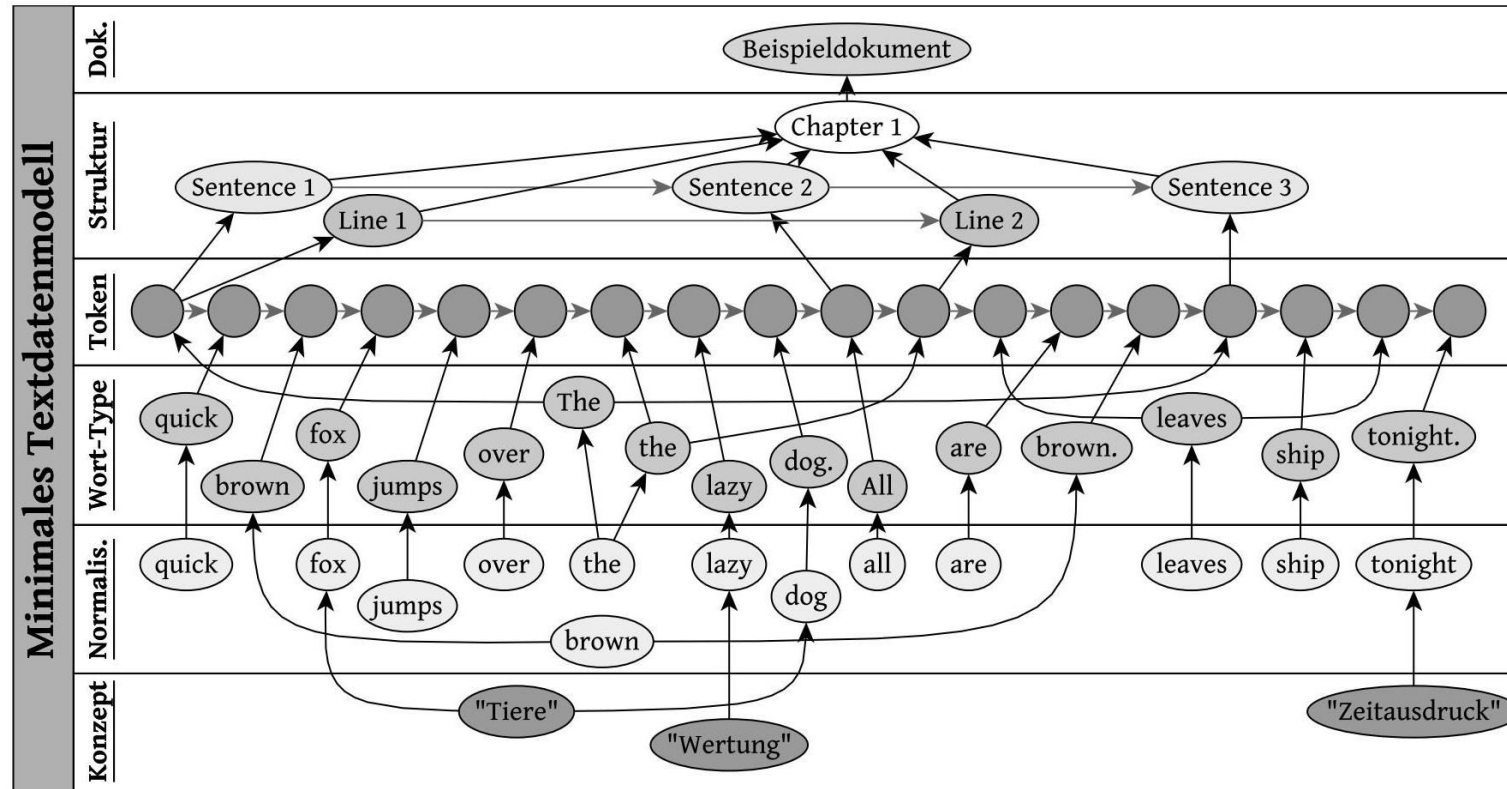
The Stemmaweb Project: Tools and techniques for empirical stemmatology. 2012–.

<https://stemmaweb.net/>



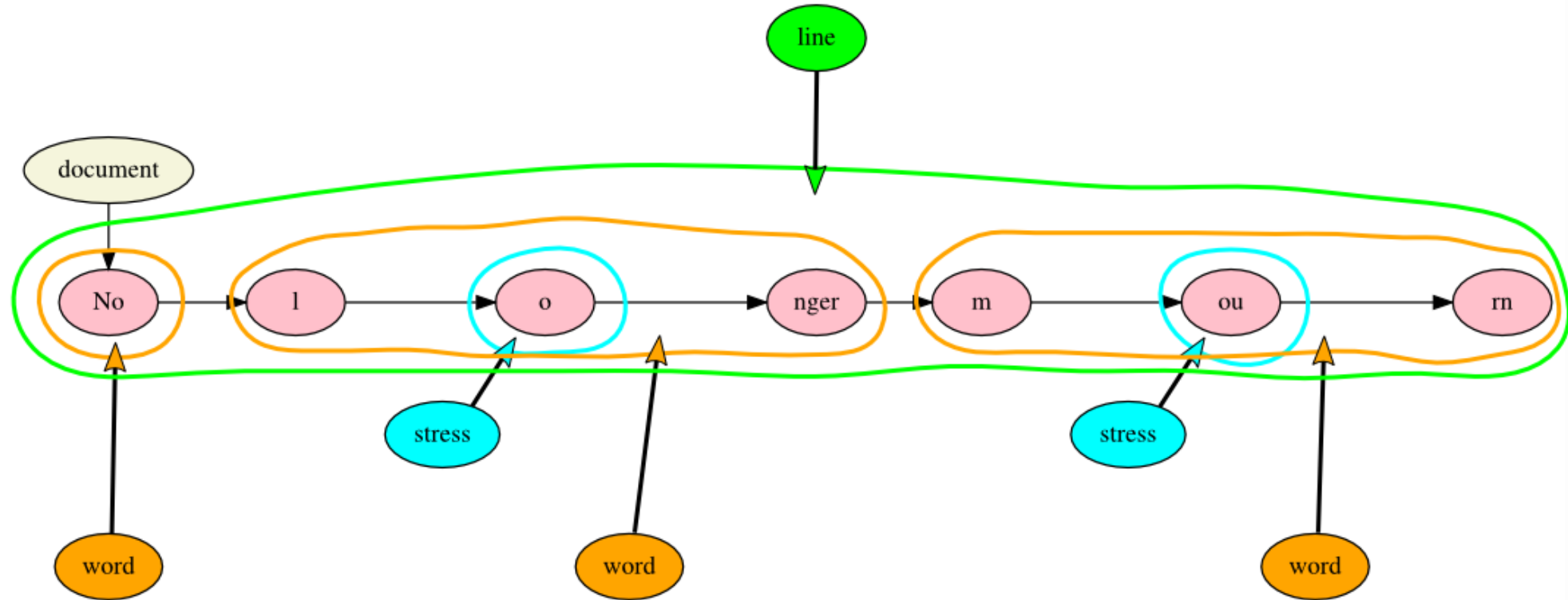
Stemmaweb: Screenshot of the Text Relationship Mapper with an excerpt from the first section of the chronicle Matthias von Edessas [The Stemmaweb Project 2012–: Screenshot)

Efer, Thomas. 2016. "Graphdatenbanken für die textorientierten e-Humanities." Dissertation, Universität Leipzig.



Kadmos: 'Schematic representation of instance data sets and links of a short example document with minimalistic text data model' [Efer 2016: 76]

Haentjens Dekker, Ronald, and David J. Birnbaum.
2017. "It's More Than Just Overlap: Text as Graph."
In Proceedings of Balisage.



TAG. Text as Graph: "A simplified poem with word tokenization" [Decker/Birnbaum 2017: Fig. 10]

Text Graph Ontology

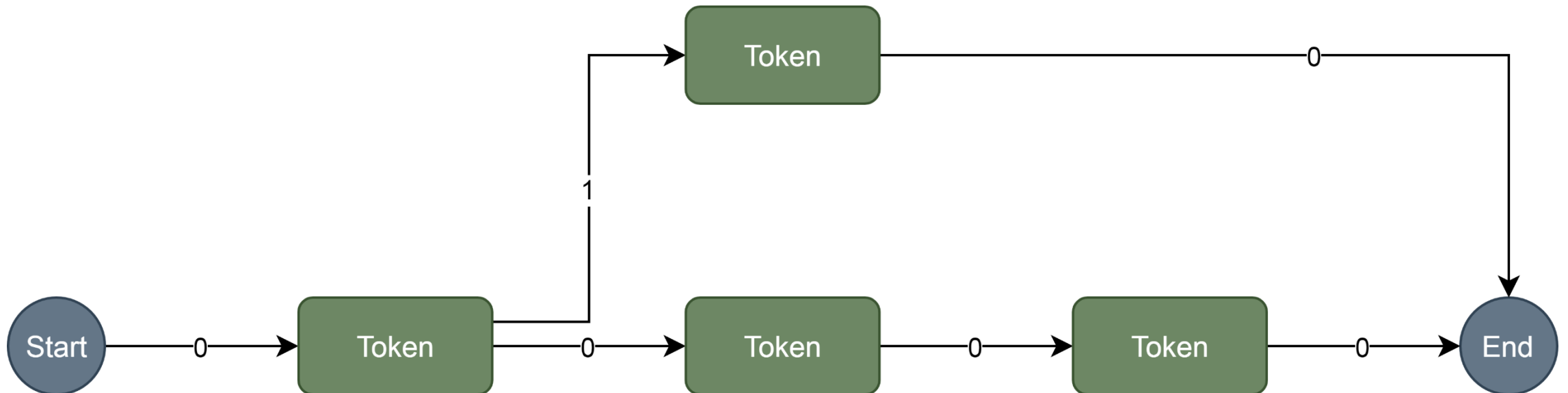
A Semantic Web approach to represent
genetic scholarly editions

Requirements

- Use of the Web Ontology Language (OWL)
- Representation of text genesis as an edge weighted directed acyclic graph
- ‘Words’ should be the basic form of segmentation
- ‘Words’ need to be further segmentable on character level
- Annotation of Text Stages and Token Groups
- Relations between Text Stages

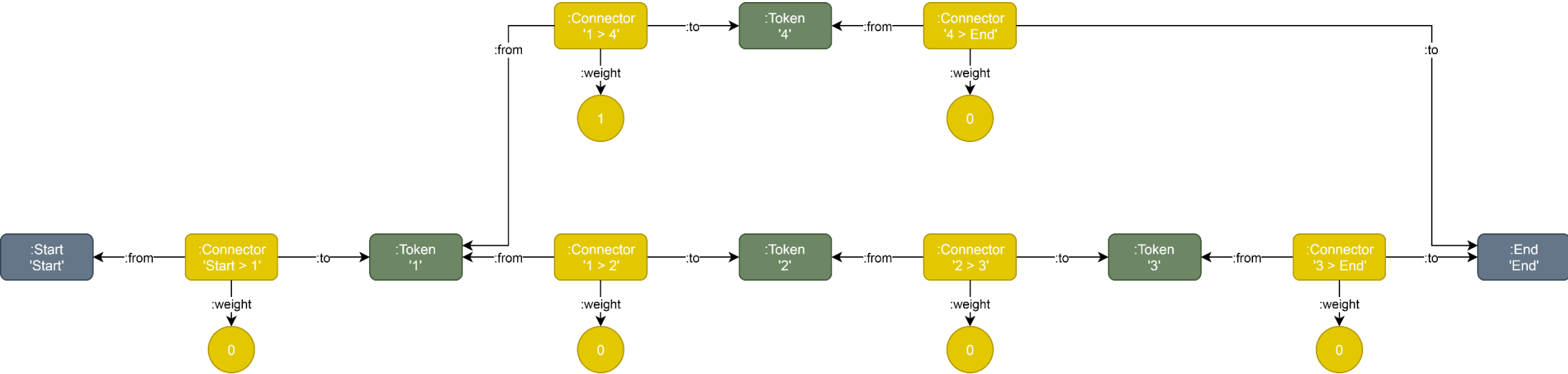
Basics

- A Text should be represented as an **acyclic directed graph**
- A **start and an end node** should exist for each text
- Variation of a text should be expressed as **weighted edges**
 - Weights allow an easy chronological reconstruction text stages



Basics as RDF solution

- Weighted Edges are not allowed because of S>P>O
 - need to be represented as individual nodes



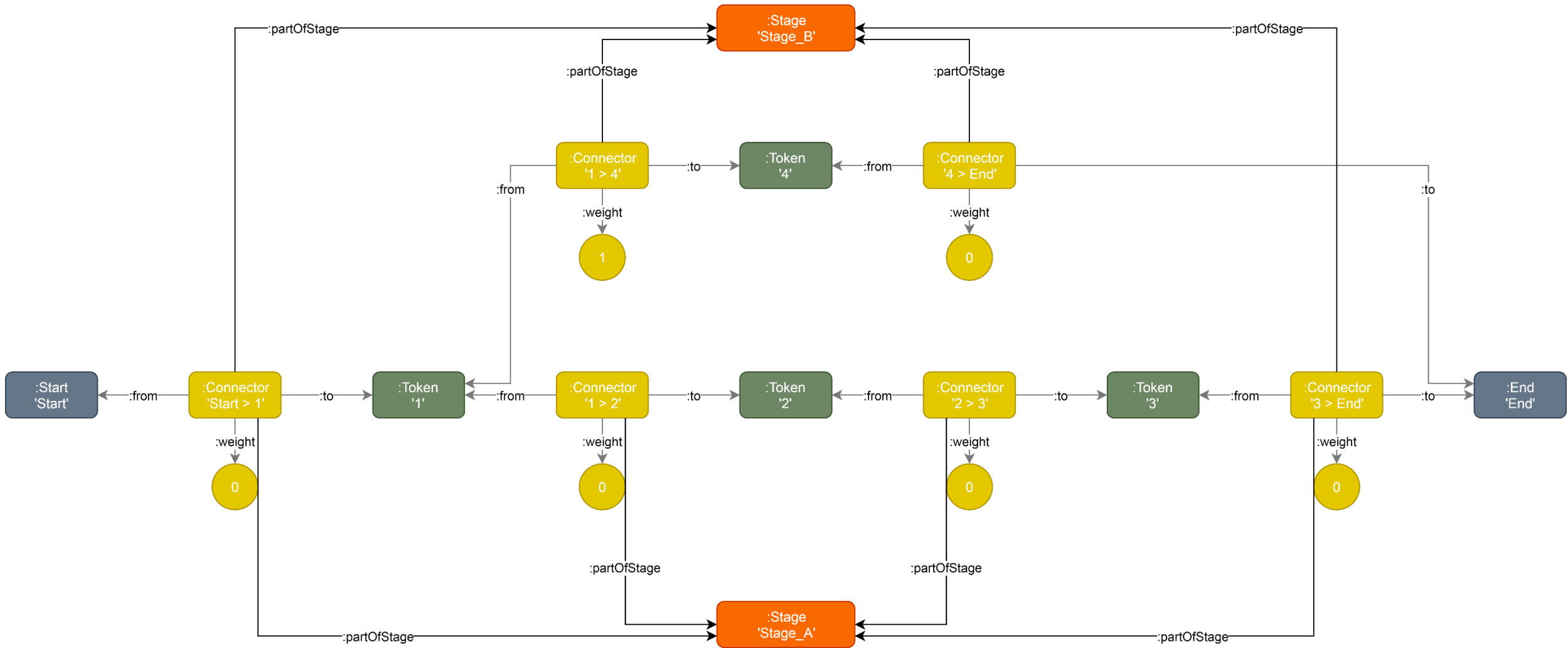
Basic Classes and Properties

Classes	Description
:Token	A 'Token' is a separated entity of a larger instance.
:Start	The start point of a graph of tokens.
:End	The end point of a graph of tokens.
:Connector	A 'Connector' represents a weighted connection between two nodes.

Object Properties	Description
:from	Domain ':Connector'; Range ':Start', ':Token'
:to	Domain ':Connector'; Range ':End', ':Token'

Data Property	Description
:weight	Domain ':Connector'; Range 'xsd:int'

Text Stages

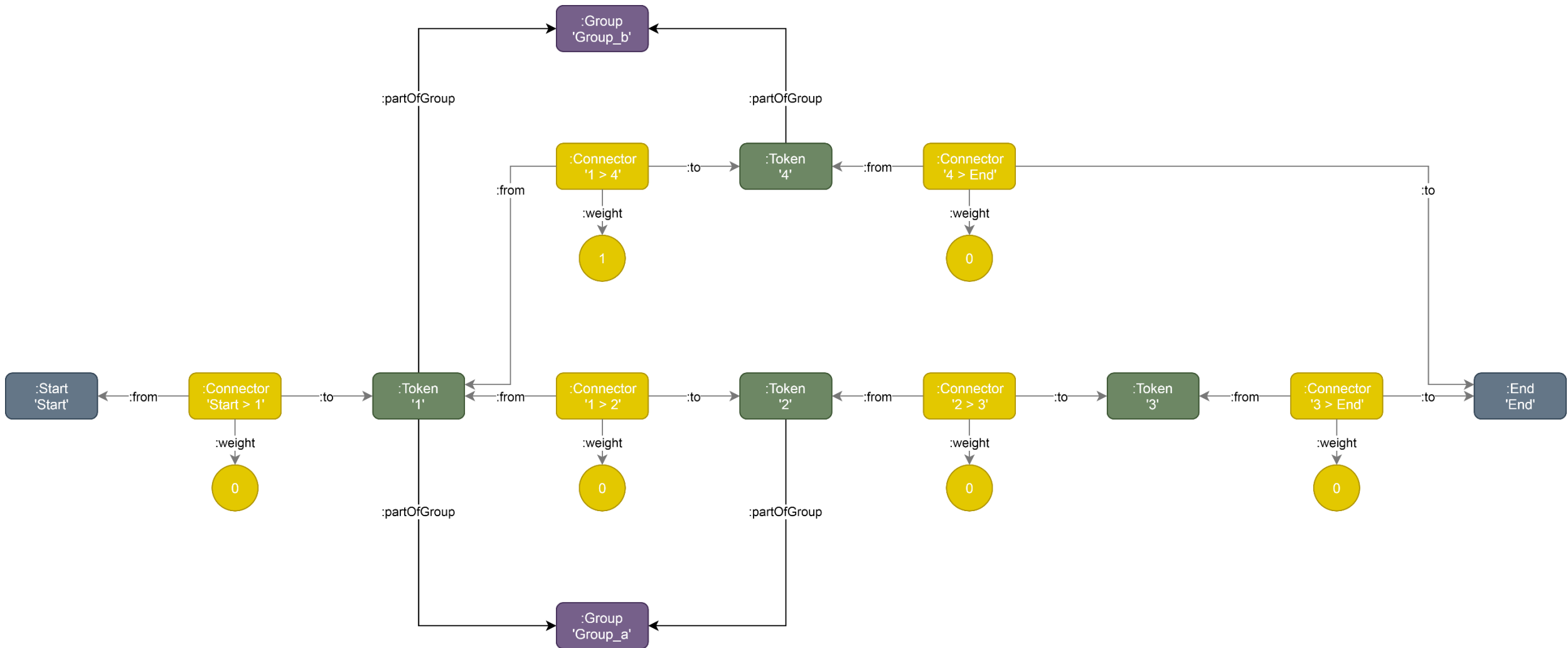


Text Stage Classes and Properties

Classes	Description
:Stage	A 'Stage' represents a stage of an instance.

Object Properties	Description
:partOfStage	Domain ':Connector'; Range ':Stage'

Annotation Groups



Annotation Groups Classes and Properties

Classes**Description**

:Group

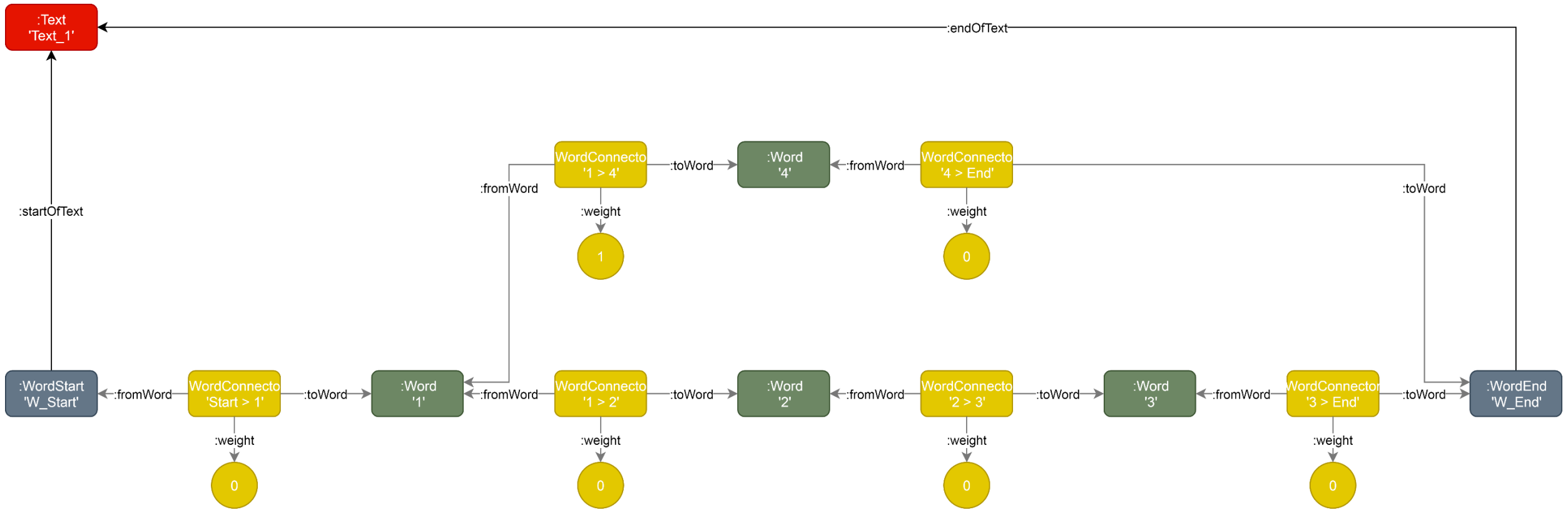
A annotation 'Group' of Tokens.

Object Properties**Description**

:partOfGroup

Domain ':Token'; Range ':Group'

Text Segmentation

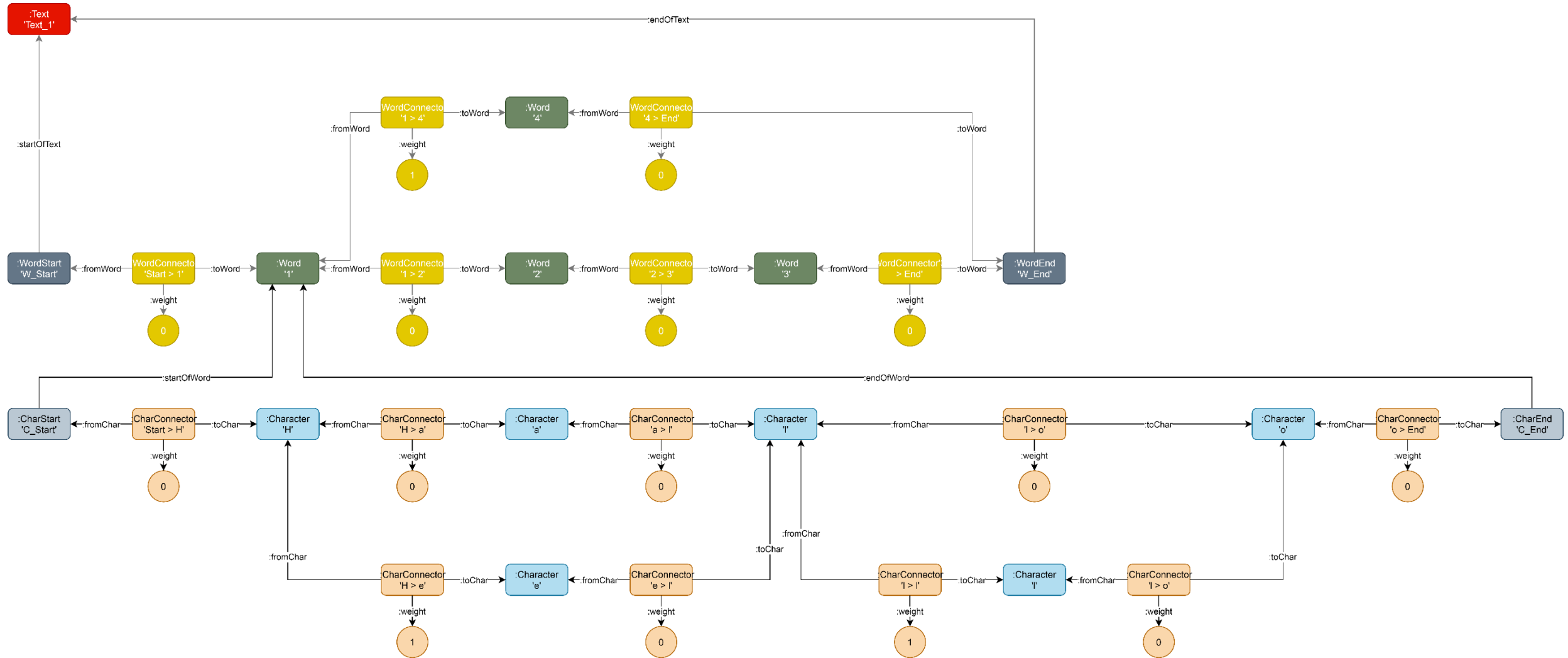


Text Segmentation

Classes	Description
:Text	Basic node for a text.
:Word	A segmented token on word level.
:WordStart	Start node on word level.
:WordEnd	End node on word level.
:WordConnector	Connector on word level.

Object Properties	Description
:fromWord	Domain ':WordConnector'; Range ':Word', ':WordStart'
:toWord	Domain ':WordConnector'; Range ':Word', ':WordEnd'
:startOfText	Domain ':WordStart'; Range ':Text'
:endOfText	Domain ':WordEnd'; Range ':Text'

Character Level



Character Level

Classes	Description
:Character	A segmented character including diacritics or abbreviations on token level
:CharacterStart	Start node on character level
:CharacterEnd	End node on character level
:CharacterConnector	Connector on character level

Object Properties	Description
:fromCharacter	Domain ':WordConnector'; Range ':Character', ':CharacterStart'
:toCharacter	Domain ':WordConnector'; Range ':Character', ':CharacterEnd'
:startOfWord	Domain ':CharacterStart'; Range ':Word'
:endOfWord	Domain ':CharacterEnd'; Range ':Word'

A real world example

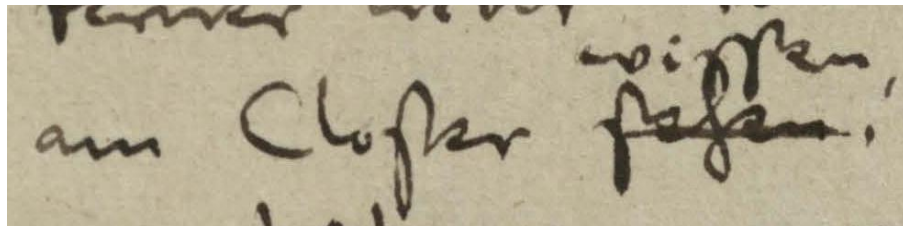
am Clofter

<subst>

<del seq="1">fehen

<add seq="2">wiſſen</add>

</subst>



```
<obj:w1> a tgo:Word ;  
  rdfs:label "am" .
```

```
<obj:w2> a tgo:Word ;  
  rdfs:label "Clofter" .
```

```
<obj:w3> a tgo:Word ;  
  rdfs:label "fehen" .
```

```
<obj:w4> a tgo:Word ;  
  rdfs:label "wiſſen" .
```

```
<obj:c1> a tgo:WordConnector ;  
  tgo:fromWord <obj:w1> ;  
  tgo:toWord <obj:w2> ;  
  tgo:weight "0" .
```

```
<obj:c2> a tgo:WordConnector ;  
  tgo:fromWord <obj:w2> ;  
  tgo:toWord <obj:w3> ;  
  tgo:weight "0" .
```

```
<obj:c3> a tgo:WordConnector ;  
  tgo:fromWord <obj:w2> ;  
  tgo:toWord <obj:w4> ;  
  tgo:weight "1" .
```

DEM N M D die 27 Novembris 1577

Durch Loth. F. C. von, vord. hoch
wird des Ruffen Kays zum Consulat
ausgesetzt auf L. & C. schreibet wegen des
Pfeuerzins Klosters neyfang und wie
schiff verfahren, und zu erregung dessel.
+ Das der Kays bey befinden, + ferner alle schiff verfahren
früheren der erwerb sie wie selber
veränderung billig Anweisung, K. P. K.
das sie auch den secht einen baron
gepferet, der des wren fanges bewandert
verantw. stellt, deshalb den studium
verfolget sein wolle des in dem orten
mit verfahren C. vord. all mit übertrag
ganz fassen mit dem bey wolle
ungleich verfahren, sondern alle in best.
L. & C. vord. leben, und der Kays die wren
knecht verfahren und sie in fesselt, und
den in bewandert aufzurufen george,
Es bein sie doch zu selber best.
woll anders gelogen sein + ferner vord.
gleich sie auch die wren george des orten
Anweisung Anweisung, oder der Kays fange
ungleich sei auf L. & C. voranston, einen
gandem in des kloster Anweisung, und
gebore auch vord. alle schiff verfahren
nicht des wren george aber den wie L. &
C. mit verfahren, das ist vord. wren
woll und wrenlicher schand, (gepfenige
den ungleich) ferner aufschreiben vord.
de, so vord. wren des kloster schiff ungleich
sien, und der selber schiff ferner george
vord.
Es vord. ist in diesen an der grade so al
vord. ist in wren vom Kloster best.
wollen eine Anweisung machen. da wren
e. f. g. Dieser Kays Anweisung. und selber Landhofen ge. stellt.

ganz fassen mit
L. & C.
woll +
wrenlicher
mittel. schiff
die ferner
der Anweisung
so viel wie be
nicht. vord.
haben ge. ferner
aber besser ungleich
erfahren ungleich.
+

woll vord. vord. f. g.
schifflicher ferner
vord. in der Kays
vord. wren selber
verfähret mit z. b.
kays ist. ferner
so auch in der ferner
vord. wren
e. f. g. Dieser Kays Anweisung. und selber Landhofen ge. stellt.



DEM N M D die 27 Novembris 1577

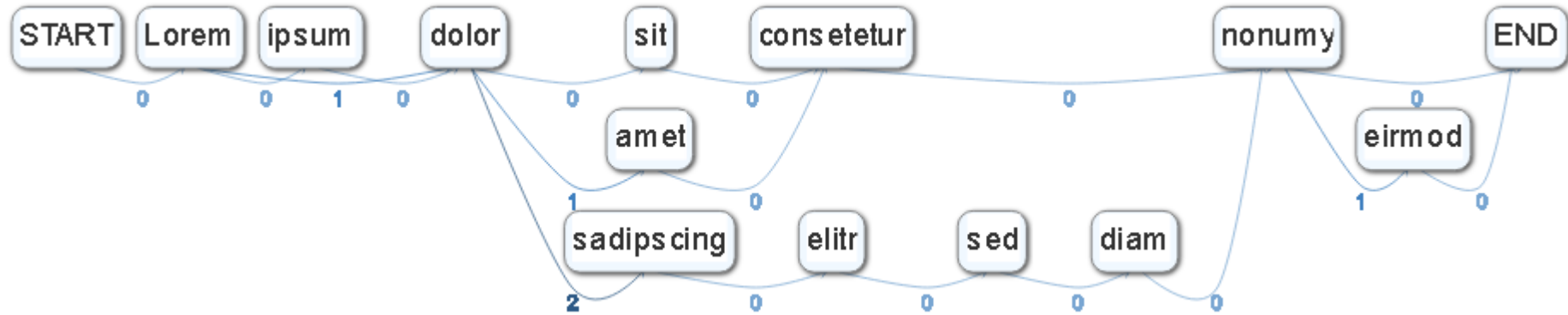
Durch Loth. F. C. von, vord. hoch
wird des Ruffen Kays zum Consulat
ausgesetzt auf L. & C. schreibet wegen des
Pfeuerzins Klosters neyfang und wie
schiff verfahren, und zu erregung dessel.
+ Das der Kays bey befinden, + ferner alle schiff verfahren
früheren der erwerb sie wie selber
veränderung billig Anweisung, K. P. K.
das sie auch den secht einen baron
gepferet, der des wren fanges bewandert
verantw. stellt, deshalb den studium
verfolget sein wolle des in dem orten
mit verfahren C. vord. all mit übertrag
ganz fassen mit dem bey wolle
ungleich verfahren, sondern alle in best.
L. & C. vord. leben, und der Kays die wren
knecht verfahren und sie in fesselt, und
den in bewandert aufzurufen george,
Es bein sie doch zu selber best.
woll anders gelogen sein + ferner vord.
gleich sie auch die wren george des orten
Anweisung Anweisung, oder der Kays fange
ungleich sei auf L. & C. voranston, einen
gandem in des kloster Anweisung, und
gebore auch vord. alle schiff verfahren
nicht des wren george aber den wie L. &
C. mit verfahren, das ist vord. wren
woll und wrenlicher schand, (gepfenige
den ungleich) ferner aufschreiben vord.
de, so vord. wren des kloster schiff ungleich
sien, und der selber schiff ferner george
vord.
Es vord. ist in diesen an der grade so al
vord. ist in wren vom Kloster best.
wollen eine Anweisung machen. da wren
e. f. g. Dieser Kays Anweisung. und selber Landhofen ge. stellt.

ganz fassen mit
L. & C.
woll +
wrenlicher
mittel. schiff
die ferner
der Anweisung
so viel wie be
nicht. vord.
haben ge. ferner
aber besser ungleich
erfahren ungleich.
+

woll vord. vord. f. g.
schifflicher ferner
vord. in der Kays
vord. wren selber
verfähret mit z. b.
kays ist. ferner
so auch in der ferner
vord. wren
e. f. g. Dieser Kays Anweisung. und selber Landhofen ge. gestellt.

Work in progress

- Converter from TEI-XML to TGO RDF and vice versa
- Viewer and editor for TGO texts



Thank you!



<https://github.com/Wolkenstein/tgo-ontology>

Bibliography

- Andrews, T. L., and C. Mace. 2013. "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation Through Graph Analysis of Texts and Stemmata." *International Journal of Human-Computer Studies* 28 (4): 504–21. doi:10.1093/llc/fqt032.
- Efer, Thomas. 2016. "Graphdatenbanken Für Die Textorientierten E-Humanities." Dissertation, Universität Leipzig.
- Haentjens Dekker, Ronald, and David J. Birnbaum. 2017. "It's More Than Just Overlap: Text as Graph." In *Proceedings of Balisage: The Markup Conference 2017*, edited by B. T. Usdin, Deborah A. Lapeyre, James D. Mason, C. M. Sperberg-McQueen, and Norman Walsh. Balisage Series on Markup Technologies: Mulberry Technologies, Inc. Rockville, Maryland.
- Ide, Nancy, and Keith Suderman. 2007. "GrAF: A Graph-Based Format for Linguistic Annotations." In *ACL 2007: The LAW*, 1–8. Madison, WI.
- Schmidt, Desmond, and Robert Colomb. 2009. "A Data Structure for Representing Multi-Version Texts Online." *International Journal of Human-Computer Studies* 67 (6): 497–514. doi:10.1016/j.ijhcs.2009.02.001.
- Sperberg-McQueen, C. M., and Claus Huitfeldt. 2004. "GODDAG: A Data Structure for Overlapping Hierarchies." In *Digital Documents: Systems and Principles*, edited by Peter King and Ethan V. Munson, 139–60. Berlin, Heidelberg: Springer.