

Annotazioni collaborative di testi storici

Angelo Mario Del Grosso

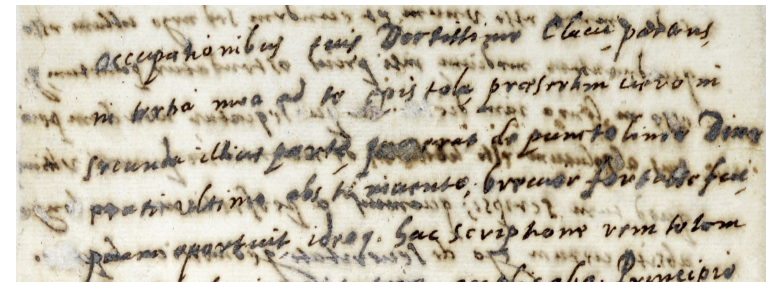
angelo.delgrosso@ilc.cnr.it

angelodel80@gmail.com



Istituto di Linguistica Computazionale

Consiglio Nazionale delle Ricerche



I relatori del Workshop 2016

❖ Angelo Mario Del Grosso, ILC-CNR

- Dottore di ricerca in Ingegneria Informatica con una tesi dal titolo “Designing a Library of Components for Textual Scholarship”. Collabora con l'**Istituto di Linguistica Computazionale di Pisa** dal 2010 all'interno della linea di ricerca orientata allo *sviluppo di componenti software per sistemi Web di linguistica e filologia computazionale* volti al trattamento di testi di tradizione medievale, a stampa e di autori moderni e contemporanei.
- Analista, progettista e sviluppatore dei servizi di elaborazione del testo.

❖ Matteo Abrate, IIT-CNR

- Dottore di ricerca in Ingegneria Informatica. Collabora con l'**Istituto di Informatica e Telematica** del CNR di Pisa dal 2010 all'interno della linea di ricerca su *Tecnologie Web e Visualizzazione Dati*.
- Progettista e sviluppatore dell'interfaccia utente.

❖ Lorenzo Mancini, ILC-CNR / APUG

- Laureato in Archivistica e Biblioteconomia, dottorando in Scienze del libro e del documento, assegnista dell'Istituto di linguistica Computazionale per il progetto Clavivus on the web.
- Annotazione delle lettere, iniziativa Clavivus@School

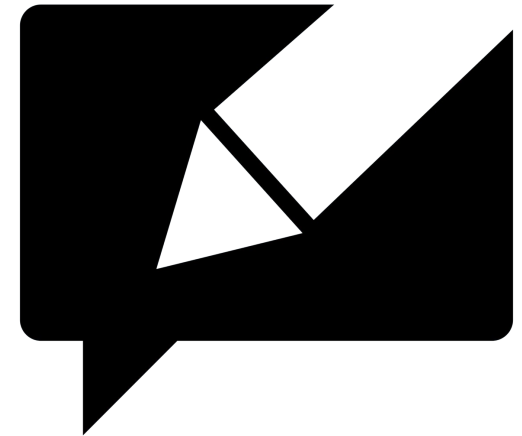
Di cosa parlerò

- ❑ Introduzione all'attività e alle procedure di annotazione
 - ❑ Architettura e Modelli concettuali
 - ❑ Annotazione e Web Semantico
- ❑ Domain Specific Languages (*DSL*)
- ❑ Il progetto **Clavius On The Web**
 - ❑ Chi è Clavius
 - ❑ Applicazione TEA, Annotarium e Omega
- ❑ Il progetto **Euforia**
 - ❑ Annotazioni **bottom-up**
- ❑ Esercitazioni Pratiche
- ❑ Conclusioni



Cosa si intende per annotazione?

- ❖ Pratica tradizionale e pervasiva per gli studiosi di documenti testuali tesa ad esplicitare e/o arricchire le informazioni di una risorsa (*diverse sfumature*):
 - indicare meta-informazioni
 - marcare il testo
 - segnare cambiamenti e/o varianti al testo
 - apporre annotazioni libere (a margine)
 - scrivere commentari su porzioni di testo
 - condividere commenti, note e riflessioni
 - mettere in relazione elementi del documento
 - mettere in relazione elementi interni della risorsa con elementi esterni
 - aggiungere (**anche automaticamente**) informazioni descrittive e analitiche
 - linguistiche, lessicali, entità nominate, tagging, classificazione ...
 - agevolare il “close” reading e/o preparare training sets

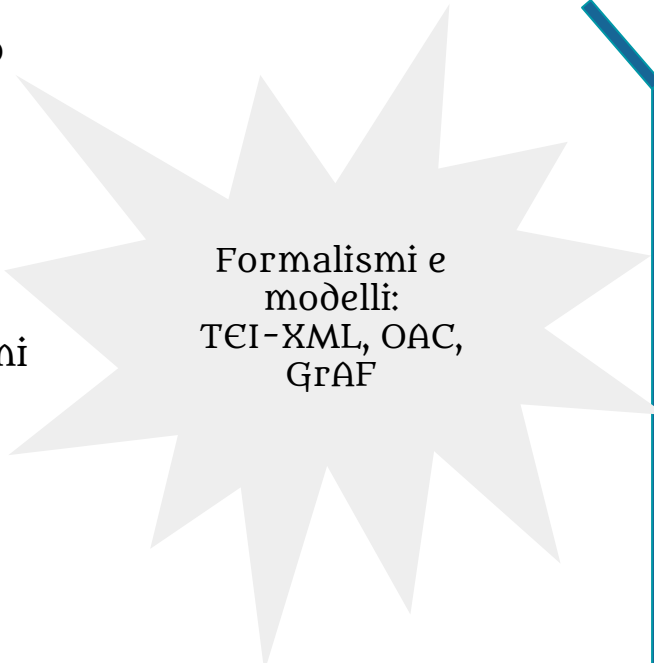


Annotazioni in line: informazioni inserite direttamente nel testo come segni di marcatura

Vantaggi: Sono facilmente gestibili da un umano e rintracciabili direttamente nel documento di origine.

Svantaggi: esplosione del documento di origine; i formalismi più comuni non gestiscono efficientemente annotazioni multidimensionali e con gerarchie sovrapposte.

[preferito nella codifica del testo]



Formalismi e modelli:
TEI-XML, OAC,
GrAF

Annotazioni in stand-off: informazioni riportate attraverso un meccanismo di puntamento alla porzione della risorsa

Vantaggi: Risorse annotabili anche senza avere il documento di origine; possibilità di gestire livelli sovrapposti in modo naturale.

Svantaggi: allineamento dei dati; ridondanza; performance.

[preferito nell'analisi dei corpora]

Testo grezzo:

ciao mondo

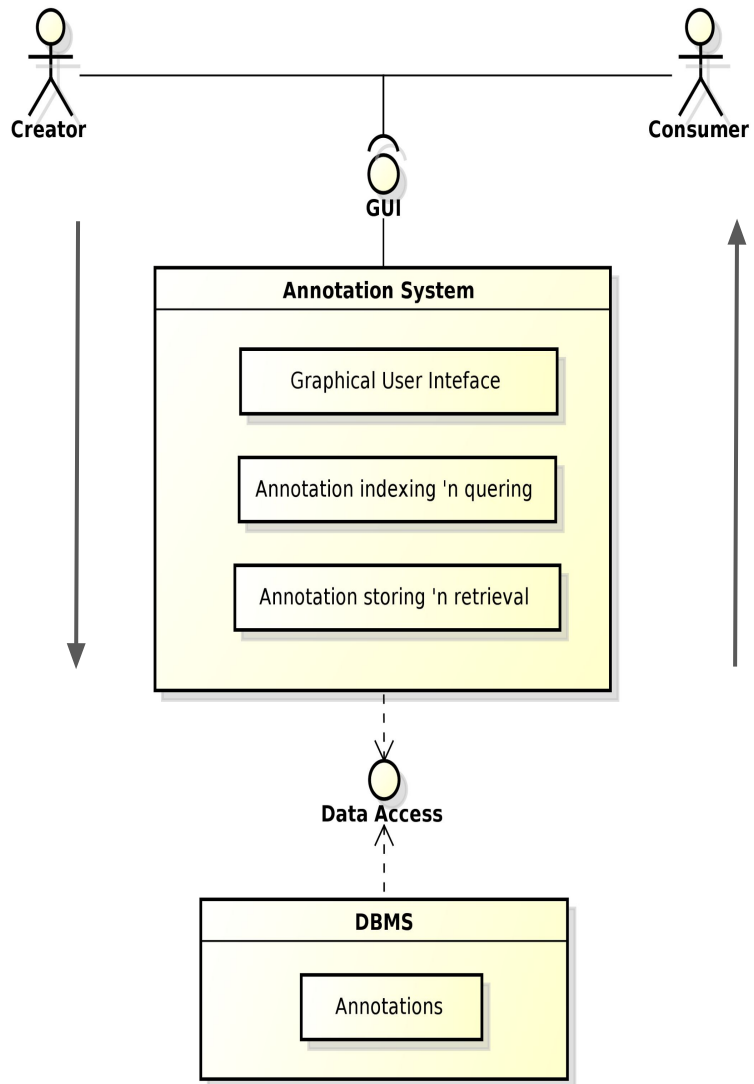
esempio inline:

```
<i>ciao
<b>mondo</b></i>
```

esempio stand-off:

```
[0-10]:italic
[5-10]:bold
```

Architettura dei sistemi di annotazione



Sistemi di annotazione:

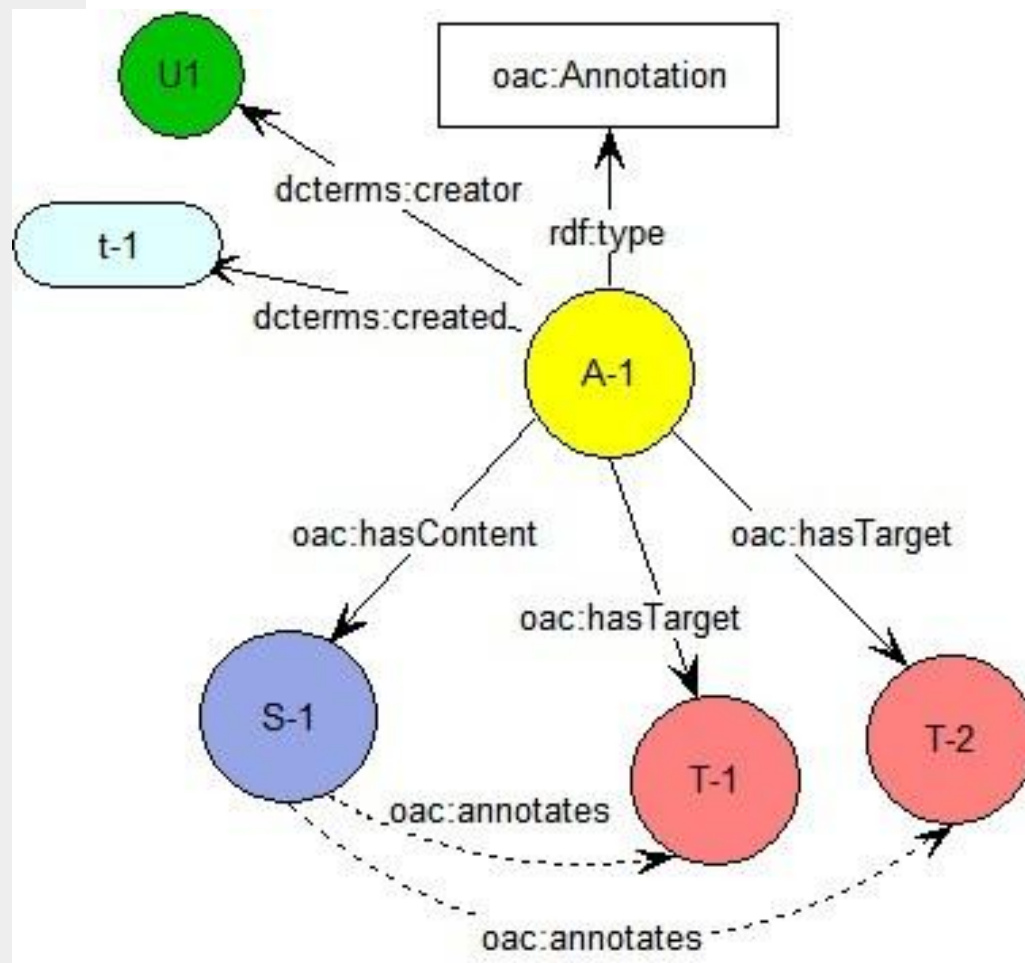
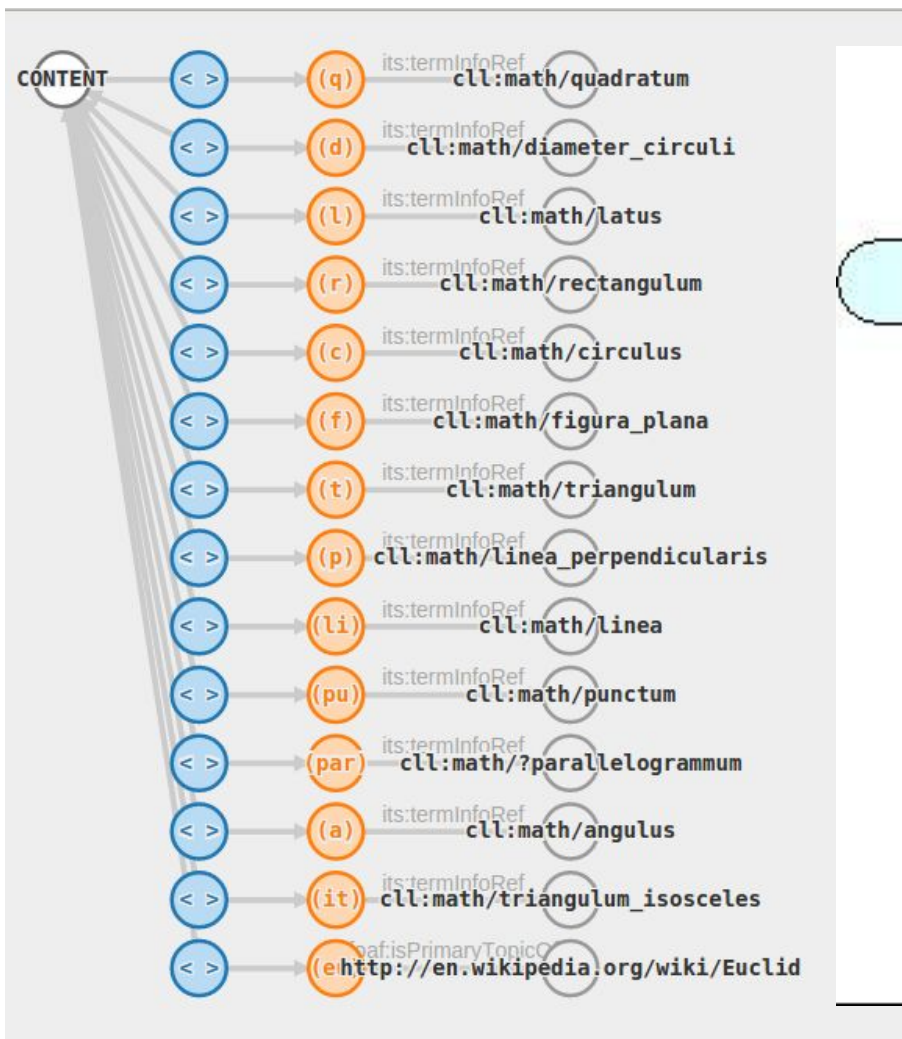
- Modulo di creazione, popolamento e collegamento
- Modulo di indicizzazione e ricerca (navigazione)
- Modulo di persistenza e recupero (storage)

Approfondimenti bibliografici:
Agosti, Hunter, Boot

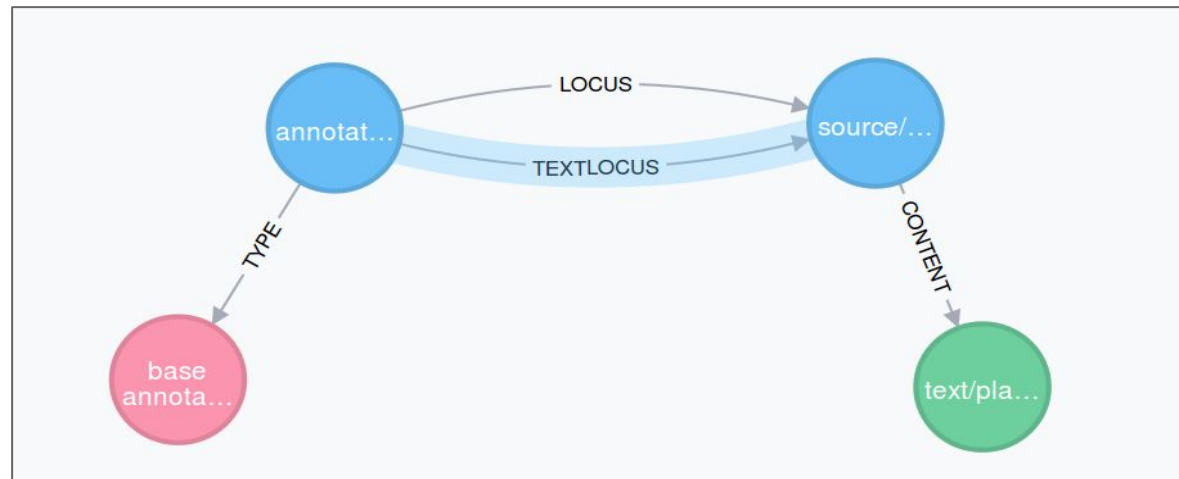
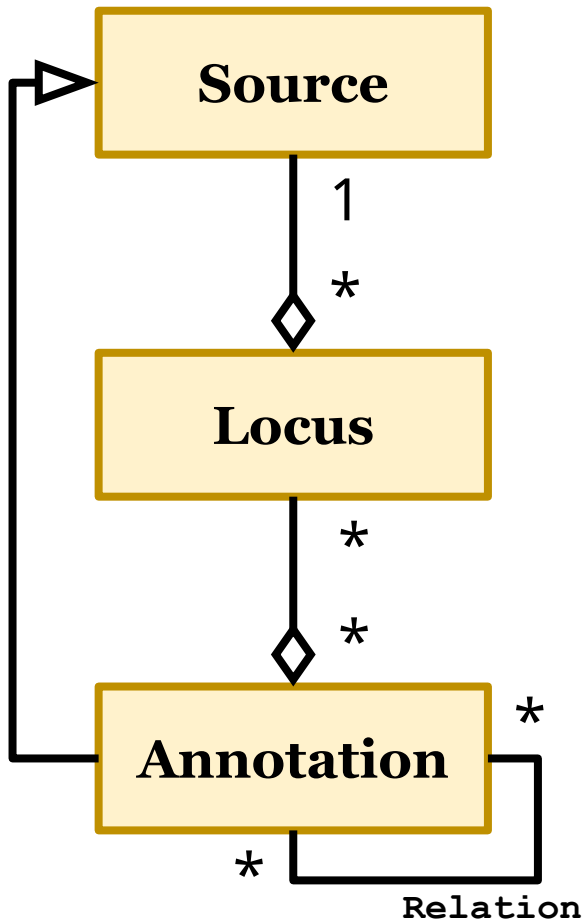
Confronto tra alcuni strumenti di annotazione

Tool Name	Web	Standoff	DSL	LOD	Free Tags	Image	Reactivity	Disc Loci	Customiz	Open Source	MultiInterpr
Euporia	✓	✓	✓		✓			✓	✓	✓	✓
Pundit	✓	✓		✓					✓	✓	
Annotation Studio	✓	✓			✓				✓	✓	
Brat	✓	✓		✓	✓			✓	✓	✓	✓
GATE TeamWare	✓	✓			✓				✓	✓	
Fast-CAT	✓	✓			✓	✓		✓			
CAT	✓	✓			✓			✓	✓		
Callisto		✓			✓			✓	✓	✓	
MMA2		✓			✓				✓	✓	
Textus	✓	✓		✓	✓			✓	✓	✓	
LORE		✓		✓				✓	✓	✓	✓
Pliny		✓			✓	✓			✓	✓	
Prism	✓	✓			✓					✓	✓
TILE	✓	✓				✓				✓	
Annotea	✓	✓		✓		✓		✓		✓	
Knowtator		✓		✓					✓	✓	

Modelli per le annotazioni del Web Semantico

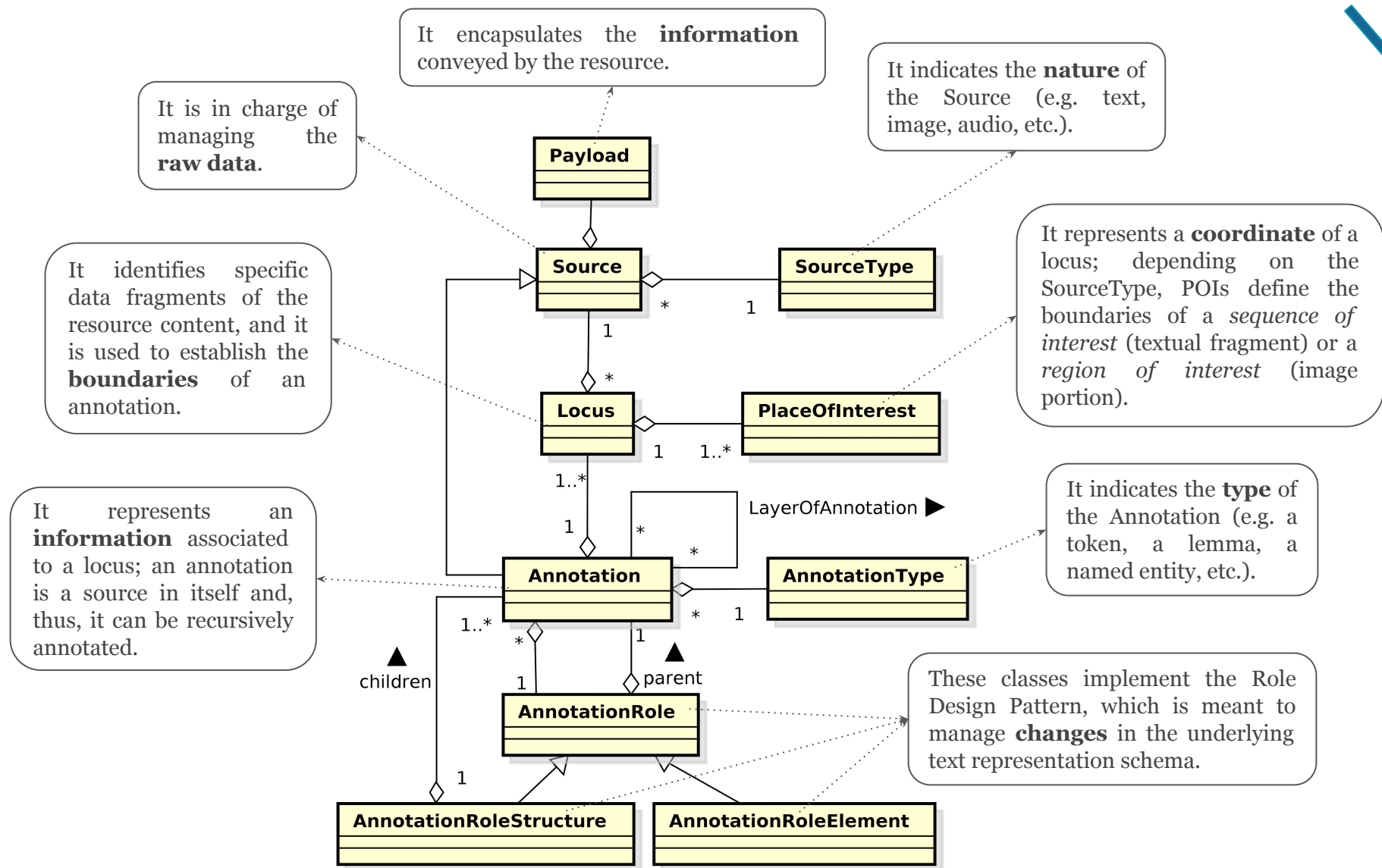


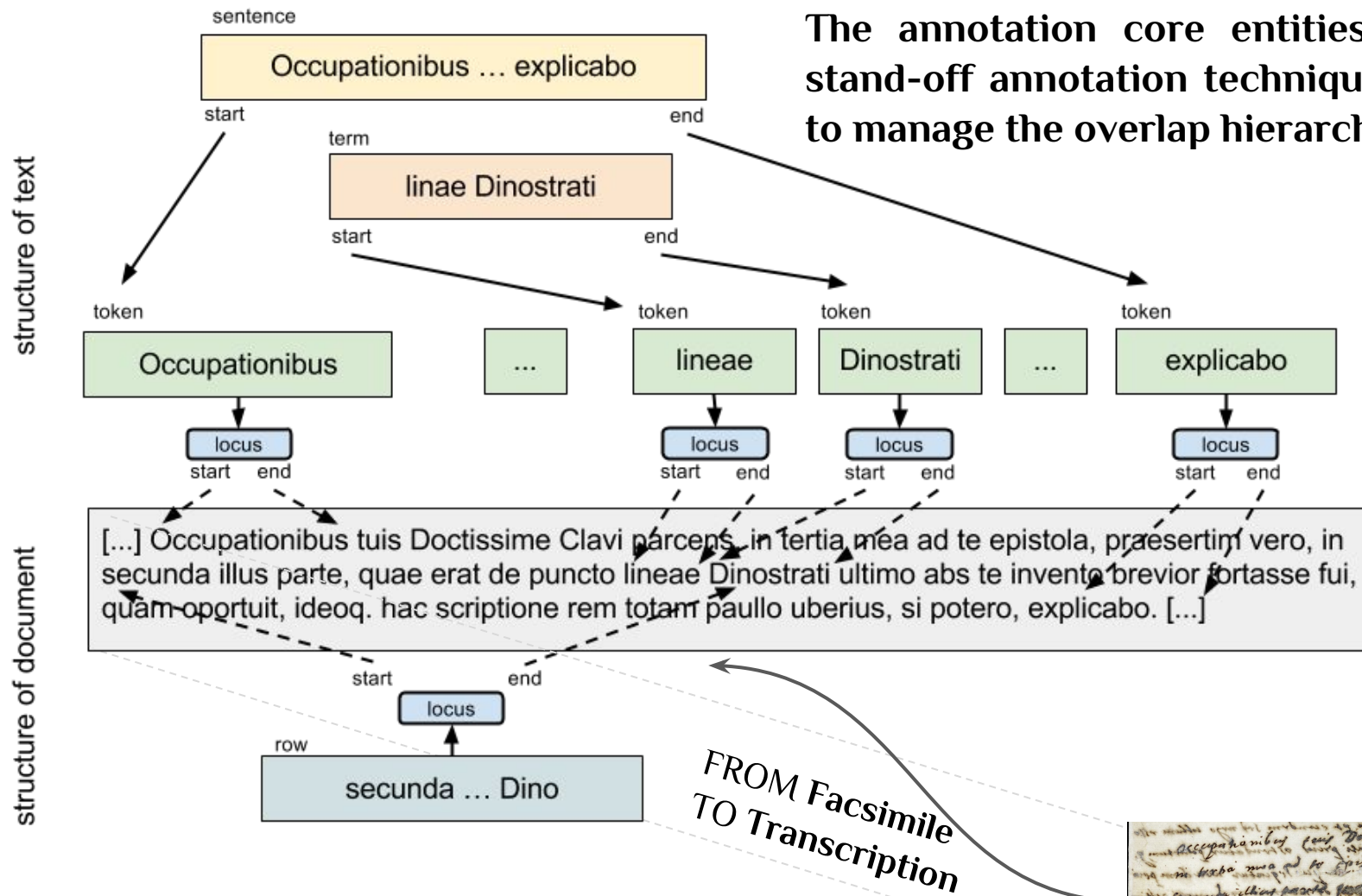
Entità Object Oriented e API



```

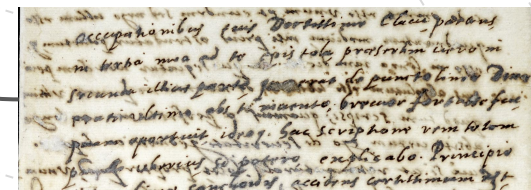
text = Text.of("Literary Text to process",
              URI.create("//source/text/000"));
annotation = AnnotationText.of("Annotation on the
text", URI.create("//annotation/text/123"));
annotation.addLocus(text, 13, 18);
annotation.save();
  
```





The annotation core entities model a stand-off annotation technique in order to manage the overlap hierarchies issue

FROM Transcription
 TO Multi-layered and multi-dimensional annotations



FROM Facsimile
 TO Transcription

Domain Specific Languages (DSL)

Definire un **linguaggio controllato** con una **sintassi** semplice, comprensibile e adatta per il **dominio** d'interesse che possa essere utilizzato come **input formale** dagli utenti e quindi allo stesso tempo “comprensibile alle macchine.

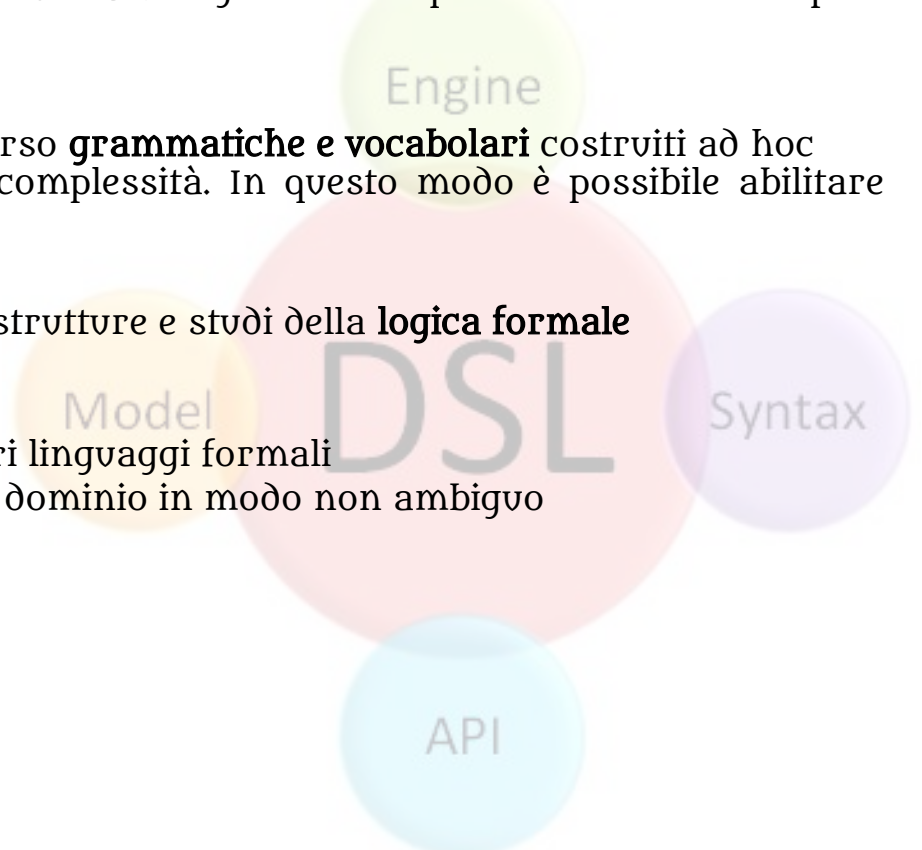
I linguaggi controllati (CNLs) sono ottenuti attraverso **grammatiche e vocabolari** costruiti ad hoc al fine di eliminare **ambiguità** e controllarne la complessità. In questo modo è possibile abilitare **sistemi automatici di elaborazione dati**

I DSL si basano su **regole linguistiche** derivanti da strutture e studi della **logica formale**

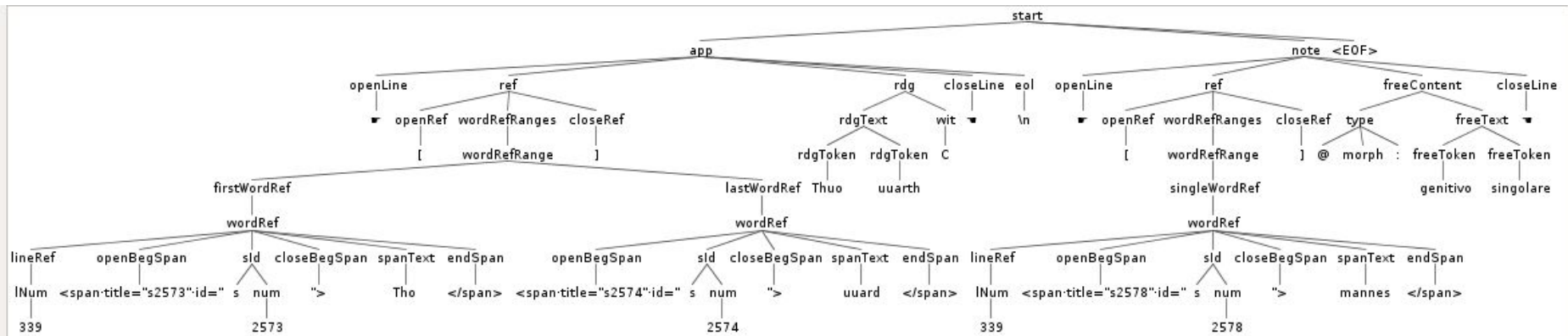
- Sintassi formale
- Semantica formale
- Possono essere mappate e trasformate in altri linguaggi formali
- adatti per rappresentare la conoscenza di un dominio in modo non ambiguo
- consistenti da un punto di vista analitico
- validabili da procedure computazionali

I vantaggi:

- Semplici da personalizzare
- Potenti da utilizzare
- Familiari per il dominio di interesse
- La curva di apprendimento è generalmente più bassa rispetto a linguaggi più vasti (per esempio TEI-XML)



Domain Specific Languages (DSL)



Un DSL può essere **formalmente** interpretato da una **context-free grammar (CFG)**. Una CFG è un insieme di **regole di riscrittura ricorsive** (productions) usate per generare pattern di stringhe.

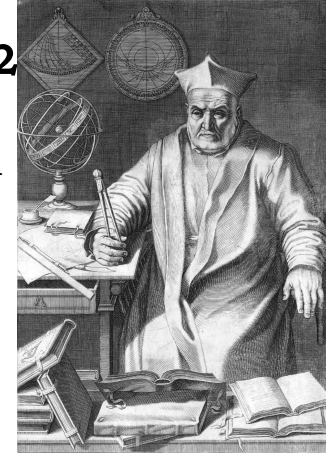
Il progetto Clavius On The Web

- costituito da **336** lettere, edite da Ugo Baldini e Pier Daniel Napolitani
- conservato principalmente nei codici **APUG 529-530** (299 lettere)
- corrispondenti da tutta Europa → lettere in latino e italiano
 - *Galileo Galilei, Tycho Brahe, Giovanni Antonio Magini, etc.*
- restauro dei codici
- argomenti: geometria, strumenti e osservazioni astronomiche, calendario
- in parte disponibili su claviusontheweb.it

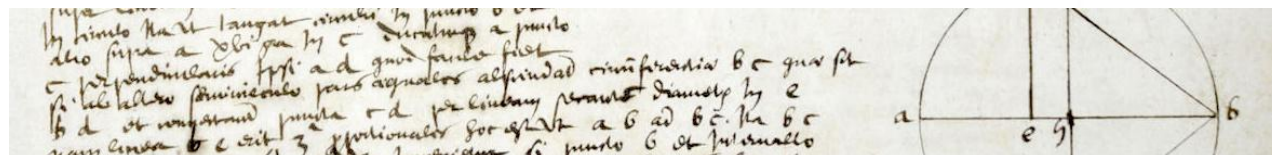


Clavius, chi?

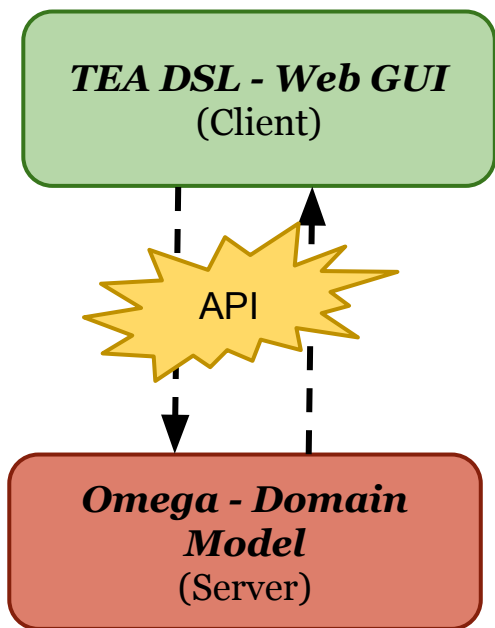
- matematico gesuita di origini tedesche (*Bamberga*): **1537-1612**
- studia a **Coimbra** e poi insegnerà a Roma per quarant'anni al **Collegio Romano**
- autorità universalmente riconosciuta del suo tempo, sebbene ancora legato alla tradizione matematica rinascimentale
- **rimformazione del calendario** (sotto *Papa Gregorio XIII* - **1582**)
- Accademia di matematica → metodo di insegnamento
- Traduzione degli *Elementi* di Euclide (**1574**)
- *Commentario De Sphaera Mundi* di Giovanni Sacrobosco (**1581**)
- non fu un innovatore



[74v]
 Admodum R. do in Christo P.
 <P. Christophoro Clavio> (Clavius)
 Societatis Iesu
 Romae.



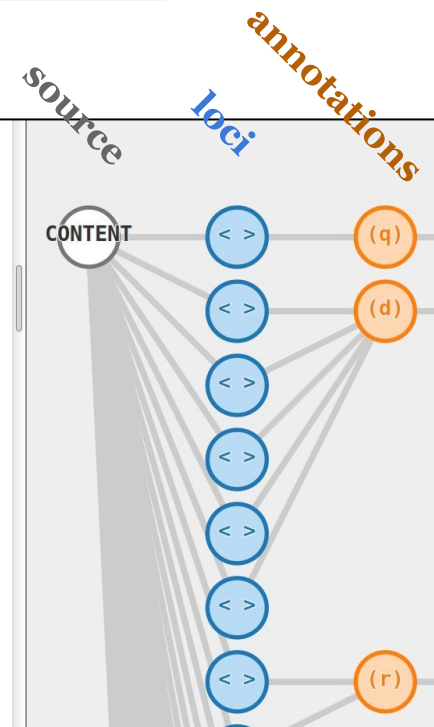
```
+++
(q) its:termInfoRef http://claviusontheweb.it/lexicon/math/quadratum
(a) its:termInfoRef http://claviusontheweb.it/lexicon/math/diameter_circuli
(idex) rdfs:comment " this is my comment "
(Clavius) owl:sameAs dbr:Christopher_Clavius
```



TEA Search Lexica

Abscindatur circumferentiae bc quae sit bd et connectantur <puncta>(p) c d per <lineam>(ln) secantem <diametrum >(d) in e nam <linea>(ln) be erit 3a proportionalis hoc est ut ab ad bc ita bc ad be per 8. 6.i 4.a vero inveniuntur si <puncto>(p) b et intervallo be sectio <circuli>(c) versus c et d fiat, donec <lineis rectis>(lnr) bc et bd occurrat in <punctis>(p) f et g quae per <lineam>(ln) fg connecti debent secantem <diametrum>(d) in h. tunc enim <linea>(ln) hb quarta proportionalis erit nam ut se habet ab. ad bc. ita bc ad be sed ut bc ad be ita bf hoc est be ad bh per 2am 6.i Igitur quemadmodum ab ad bc. Ita bc ad be et be ad bh eademque arte inveniuntur 5.a proportionalis si <puncto>(p) b et intervallo bh portio <circuli>(c) describatur quae secet <lineas>(ln) cb. Et bnd et <puncta>(p) sectionum iungantur per <lineam>(ln) secantem <diametrum>(d) interiecta enim inter hanc <lineam>(ln) et <punctum >(p)b. 5a erit proportionalis. Hinc constat datis quibuscumque <linei>(ln) facile inveniri posse duas medias continuas proportionales si hoc problema inventum esset datis duabus <lineis >(ln)//

[73r]
 1. a Datis differentiis diametri rectanguli a lateribus rectum ambientibus et latera et diametrum invenire



Strumenti

- Trascrizione e Annotazione: [TEA \(Text Encoder and Annotator\)](#)
- Ricerca e navigazione: [Annotarium](#)
- Indicizzazione e retrieval: [Omega-ClavusWeb](#)
- [Clavus Annotation](#)
- [Clavus Knowledge tree](#)
- [Clavus Search](#)
- [Clavus Visualization](#)
- [Clavus linguistic Analysis](#)
- [Clavus End Point for Linked Open Data](#)
- [Navigazione Knowledge Graph](#)



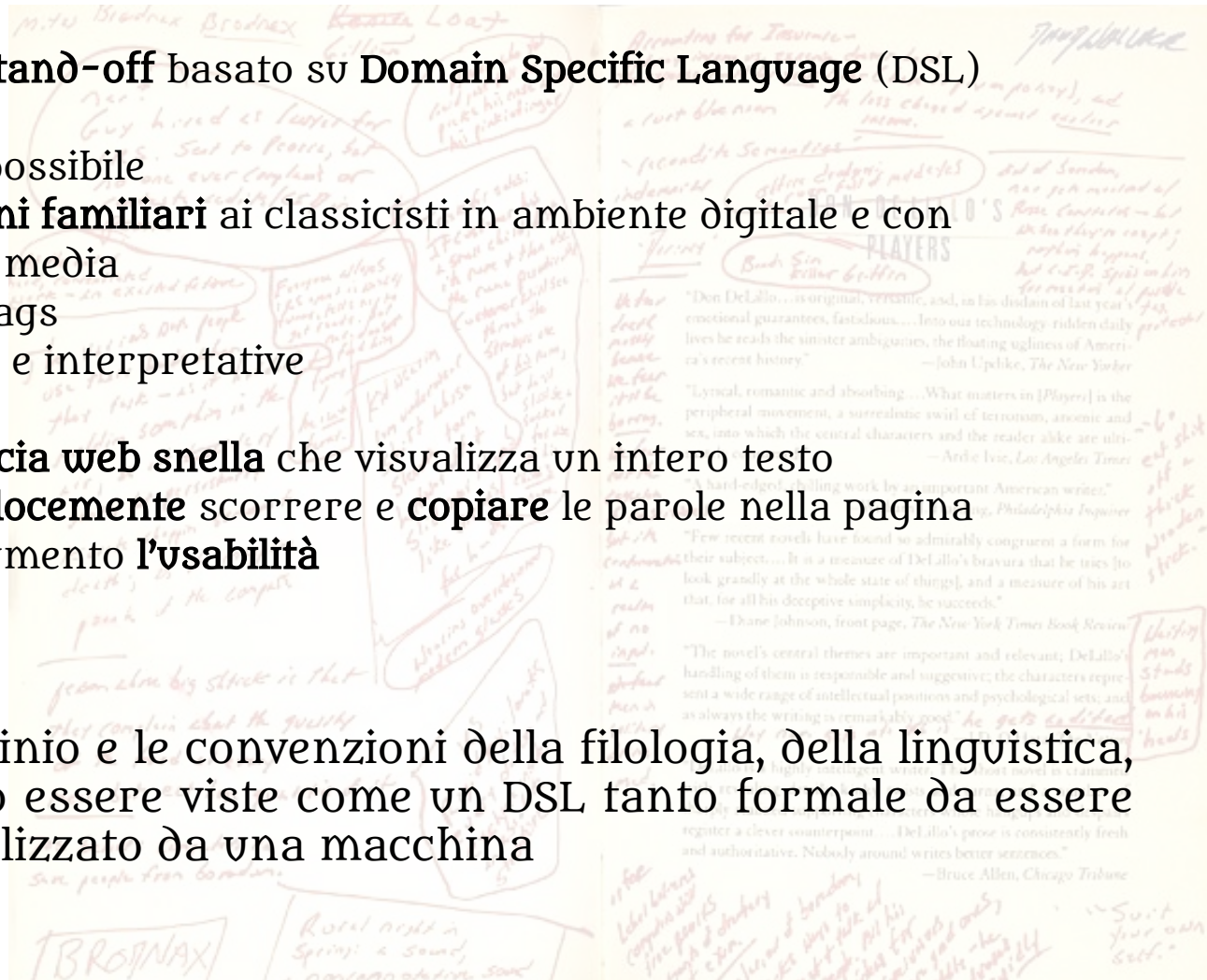
Clavius
on the Web

Euforia: Annotazione stand-off con DSL

Sistema di annotazione **stand-off** basato su **Domain Specific Language (DSL)**

- Quanto più **conciso** possibile
- Basato su **convenzioni familiari** ai classicisti in ambiente digitale e con riferimento ai social media
 - **Citazioni**, hashtags
 - **Varianti** testuali e interpretative
- Dotato di **un'interfaccia web snella** che visualizza un intero testo
 - L'utente deve **velocemente** scorrere e **copiare** le parole nella pagina
 - La **semplicità** aumento l'**usabilità**

Il linguaggio di dominio e le convenzioni della filologia, della linguistica, della storia possono essere viste come un DSL tanto formale da essere comprensibile e analizzato da una macchina



Ευπορία: Annotazione di rituali nella tragedia di Eschilo - Agamennone

cophilab.ilc.cnr.it:8080/euporiaweb/

Most Visited Getting Started

Χορός
 228 λιτὰς δὲ καὶ κληδόνας πατρώους
 229 παρ' οὐδὲν αἰῶ τε παρθένειον
 230 ἔθεντο φιλόμαχοι βραβῆς.
 231 φράσεν δ' ἄόζοις πατὴρ μετ' εὐχὰν
 232 δίκαν χιμαίρας ὑπερθε βωμοῦ
 233 πέπλοισι περιπετῆ παντὶ θυμῷ προνωπῆ
 235 λαβεῖν ἀέρδην, στόματός
 236 τε καλλιπρώρου φυλακᾶ κατασχεῖν
 237 φθόγγον ἀραῖον οἴκοις,

- [228 λιτὰς... 249 ἄκραντοι] #h #sacrificium #hominem_sacrificare ▶
- [228 λιτὰς... πατρώους] #supplicatio #preces #lissomai ▶
- [229 παρθένειον] #virgo #victima ▶
- [231 μετ' εὐχὰν] #ritus_tempus #precatio #euche ▶
- [231 ἄόζοις] #minister ▶
- [232 δίκαν... ὑπερθε βωμοῦ_235 λαβεῖν ἀέρδην] #victimam_tollere ▶
- [232 βωμοῦ] #altaria ▶
- [232 δίκαν χιμαίρας] #capra #virgo_sicut_victima #aetas▶
- [233 πέπλοισι περιπετῆ] @vi:233_1 #victimam_vincire #vestis ▶
- [233 προνωπῆ] {@vi:233_1} #pronus ▶
- [233 προνωπῆ] @vi:233_2 #animo_relictus Medda2012 ▶
- [233 πέπλοισι... προνωπῆ] @vi:233_3 #supplicatio Bonanno2006 ▶
- [233 πέπλοισι περιπετῆ] {@vi:233_3} #vestem_tangere ▶
- [233 προνωπῆ] {@vi:233_3} #ad_genua_accidere ▶

Ευπορία: Ricerca nella tragedia di Eschilo - Agamennone

sacrificium

hominem_sacrificare

0

0

0

0

Search

A.Ag. 70 ἀπύρων - 70 ἱερῶν

A.Ag. 151 θυσίαν - 151 ἄδαιτον

A.Ag. 151 θυσίαν - 151 θυσίαν

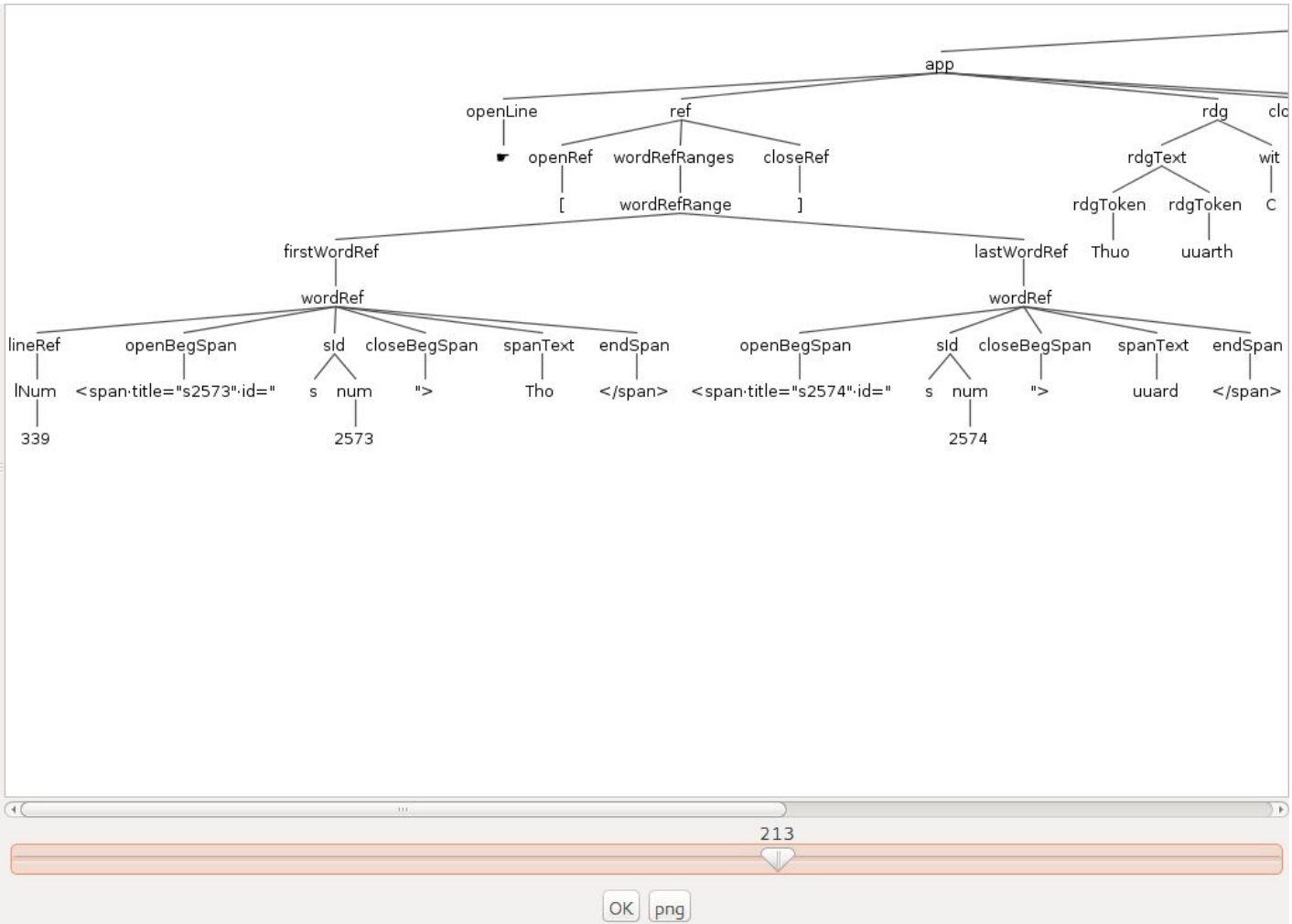
A.Ag. 228 λιτὰς - 249 ἄκραντοι.

scaenica A.Ag. 1 θεοὺς - 1673 <καλῶς>.**domus** A.Ag. 1 θεοὺς - 1673 <καλῶς>.**hiera** A.Ag. 228 λιτὰς - 249 ἄκραντοι.**sacrificium** A.Ag. 228 λιτὰς - 249 ἄκραντοι.**hominem_sacrificare** A.Ag. 228 λιτὰς - 249 ἄκραντοι.**supplicatio** A.Ag. 228 λιτὰς - 228 πατρώους**preces** A.Ag. 228 λιτὰς - 228 πατρώους**lissomai** A.Ag. 228 λιτὰς - 228 πατρώους**virgo** A.Ag. 229 παρθένειον - 229 παρθένειον**victima** A.Ag. 229 παρθένειον - 229 παρθένειον**ritus_tempus** A.Ag. 231 μετ' - 231 εὐχὰν**precatio** A.Ag. 231 μετ' - 231 εὐχὰν**euche** A.Ag. 231 μετ' - 231 εὐχὰν**minister** A.Ag. 231 ἀόζοις - 231 ἀόζοις**victimam_tollere** A.Ag. 232 δίκαν - 232 βωμοῦ**victimam_tollere** A.Ag. 235 λαβεῖν - 235 ἀέρδην,**altaria** A.Ag. 232 βωμοῦ - 232 βωμοῦ**capra** A.Ag. 232 δίκαν - 232 χιμαίρας**victima** A.Ag. 232 δίκαν - 232 χιμαίρας**homo_sicut_victima** A.Ag. 232 δίκαν - 232 χιμαίρας**aetas** A.Ag. 232 δίκαν - 232 χιμαίρας**victimam_vincire** A.Ag. 233 πέπλοισι - 233 περιπετῇ**vestis** A.Ag. 233 πέπλοισι - 233 περιπετῇ**pronus** A.Ag. 233 προνωπῇ - 233 προνωπῇ**animo_relictus** A.Ag. 233 προνωπῇ - 233 προνωπῇ**supplicatio** A.Ag. 233 πέπλοισι - 233 προνωπῇ**vestem_tangere** A.Ag. 233 πέπλοισι - 233 περιπετῇ**ad_gerens_casidum** A.Ag. 233 προνωπῇ - 233 προνωπῇ

Euforia: La Grammatica formale di riferimento

```

  start
  app
  openLine
  ref
  openRef
  wordRefRanges
  wordRefRange
  firstWordRef
  wordRef
  lineRef
  lNum
  339
  openBegSpan
  <span:title="s2573".id="
  sld
  s
  num
  2573
  closeBegSpan
  ">
  spanText
  Tho
  endSpan
  lastWordRef
  closeRef
  rdq
  rdqText
  rdqToken
  Thuo
  rdqToken
  uuarth
  wit
  closeLine
  eol
  \n
  note
  <EOF>
  
```



Euporia: Esercizio - Poema antico sassone Heliand

doc http://www.h...eu/heliand/ (1 unread) - federico... doc JSP Page Preferences

cophilab.ilc.cnr.it:8080/euporiaHeliand/index.jsp?work=5

Most Visited Getting Started

Heliand. V

339 Tho uuard fon Rumuburg | rikes mannes 20
 340 obar alla thesa irminthiod | Octauianas
 341 ban endi bodskepi | obar thea is bredon giuuald
 342 cuman fon them kesure | cuningo giuhuilicun,
 343 hemsitteandiun, | so uuido so is heritogon
 344 obar al that landskepi | liudio giuueldun.
 345 Hiet man that alla thea elilendiun man | iro odil
 (6a) sohtin, 11, 1
 346 helidos iro handmahal | angegen iro herron
 bodon,
 347 quami te them cnosla gihue | thanan he cunneas
 uuas,
 348 giboran fon them burgiun. | That gibod uuarð
 gilestid
 349 obar thesa uuidon uuerold. | Uuerod samnoda
 350 te allaro burgeo gihuuem. | Forun thea bodon
 obar all 5
 351 thea fon them kesura | cumana uuarun,

◀ [339 Tho uuard] Thuo uuarth C ▶

Save

Show HTML

Show Parsing Result

352 bokspaha uueros, | endi an bref scribun
 353 suiðo niudlico | namono giuhuilican,
 354 ia land ia liudi, | that im ni mahti alettean man

◀ [] ▶

Saved

Show HTML

Show Parsing Result

Annotazioni collaborative di testi storici

Grazie!



Istituto di Linguistica Computazionale

Consiglio Nazionale delle Ricerche

Angelo Mario Del Grosso

angelo.delgrosso@ilc.cnr.it

angelodel80@gmail.com

