



Consiglio Nazionale delle Ricerche



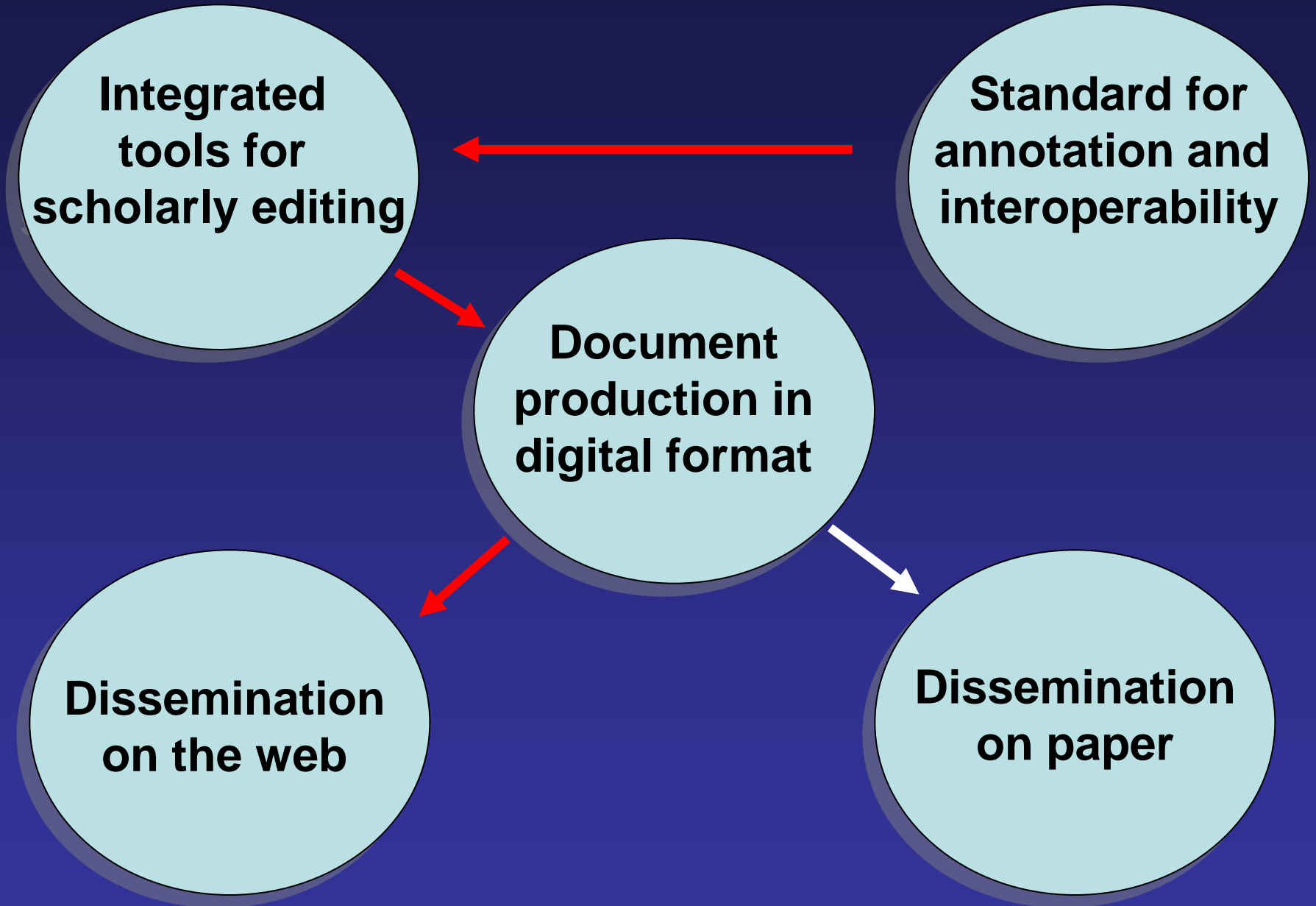
**Istituto di Linguistica Computazionale – Pisa**

**Andrea Bozzi**

# **Computer-assisted scholarly editing of manuscript sources**

Changing Publication Cultures in the Humanities  
Budapest, 25-28 november 2009

# A Model for Digital Philology



# Basic tools for scholarly editing digital documents

Text indexing and concordances

Image enhancement

Texts and images: integrated *open source* environment to manage images and texts

Annotations and variants: collaborative scholarly editing

3D: stemmatology (graphical representation of relations among witnesses)

NLP: lemmatization, morphological analysis, creation of data banks of syntactic structures, sense extraction, identification of *named entities*, ....

# Pinakes Text (PKT): the text module of Pinakes

([http://pinakes.imss.fi.it/index.php/Main\\_Page](http://pinakes.imss.fi.it/index.php/Main_Page))

- To browse digital libraries and view documents
- To edit documents (add, modify, delete)
  - For each page of a document you can specify the corresponding scanned image and text transcription
- To search documents
  - Base search: by title, by author, by volume, etc...
  - Advanced search: by text, lemmas and concepts contained in the document
- To enrich documents with meta-data (annotations) relevant for philological analysis



# Pinakes Text – Search documents by content

Users can specify one or more words to be retrieved in the documents archive

- For each word it is possible to specify
  - Type: whole word, fragmented word, lemma, ...
  - What to consider: case sensitive, stress sensitive, ...
  - Where to search: body, notes, titles, prose, ...
- Terms can be combined to obtain a more complex search expression
  - Boolean operators: *and*, *or*, *not*

# Pinakes Text – Documents Annotations (view)

The screenshot displays the Pinakes Text web application interface. At the top, the browser address bar shows the URL `http://localhost:8080/PkSelectionInt/annotate.jsf`. The page title is "Pinakes Text – Document visualization".

The main header features the "pinakes text" logo in green. To the right, there are logos for "Fondazione RINASCIMENTO digitale" and "CAIR". Below the header, a navigation bar contains buttons for "Volume Index", "Search", "Working Set", and "Edit Documents".

The left sidebar contains a "Document Index" section and an "Annotations" section. The "Annotations" section includes buttons for "New Annotation", "Save Annotations", and "Delete Annotation". Below these buttons is a table with columns for "Type", "Individual", and "Properties".

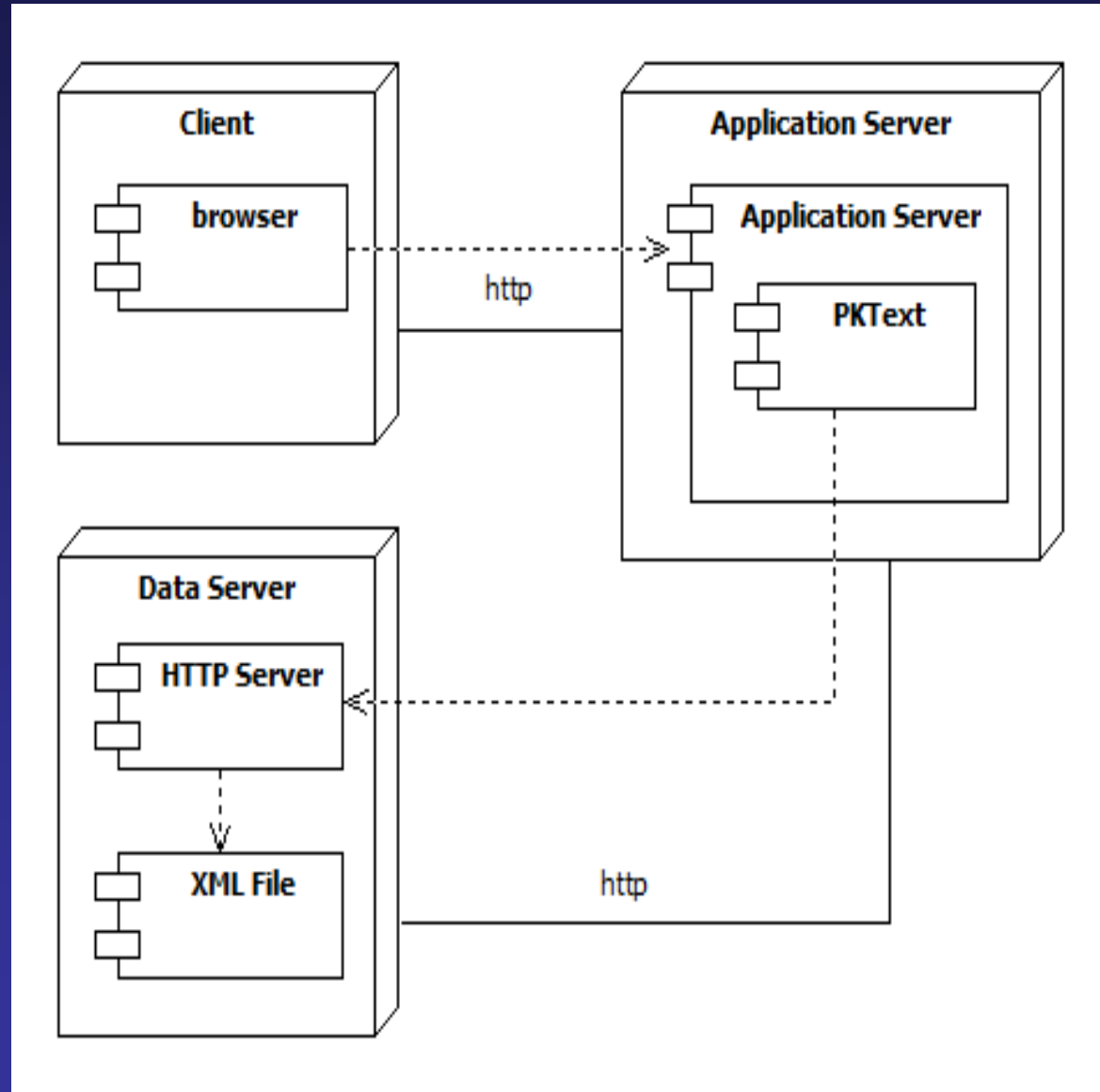
The main content area shows a document viewer. At the top of the viewer, it indicates "Volume1 page 1 of 1" and "Page numbering: Original". The document text is displayed in a monospaced font. A green highlight is visible on the text "Un po' di testo" in the left pane. In the right pane, there is a yellow highlight on the word "AVVERTIMENTO." followed by text: "meridi dal 3 agosto al 15 ottobre (pag. 656- n esame per il secondo che abbiamo riportat e che induce a nuove correzioni, le quali si altri. Con gli elementi corretti sono calcolate (pag. 665-666), e nuovamente mutano al 20 otto".

At the bottom of the viewer, it again shows "Volume1 page 1 of 1".

The status bar at the bottom left of the browser window reads "Trasferimento dati da localhost...".

# Pinakes Text - Architecture

- Web based application
- Used Tools
  - Java
  - Java Server Faces, JavalceFaces
  - Facelets
  - Lucene
  - OWL





# **Flexibility of PKT: some case studies and areas of application**

- **Greek papirology and classical philology**
- **Romance Philology**
- **Philology of ancient printed books**
- **Linguistic tools: morphological analyzer and lemmatization engine**

# Annotations and critical apparatus

The interface displays a digital image of a papyrus fragment on the left and its transcription on the right. The transcription is as follows:

1 Ἰουλίῳ Ἀσκληπιείῳ στρ(ατηγῶ)  
παρὰ Θεώνος τοῦ Ἀρτάλου ἰατροῦ  
τῶν ἀπ' Ὀξυρυγχῶν πόλεως. Τῇ ἐβ-  
δόμῃ τοῦ ἐνεστῶτος μηνὸς Σωτη-  
5 [ρείου] ἐπετράπη μοι ὑπὸ Πτολεμαίου  
Ἰ [υ]πηρετοῦ ἐφιδεῖν  
Ἰ [ανος ἀπὸ συμπλο-  
Ἰ κῆς ἰατος ὁδοὺς ἐξε  
Ἰ [σκεψάμενος οὖν  
10 [τῇ αὐτῇ ἡμέρᾳ ἐπακολουθοῦν-  
[τος] τοῦ αὐτοῦ ὑπηρετοῦ ἐγνω  
[τ]οῦτον ἐξηρ.μενον οὖν  
Ἰ [ .εσκα..] Ἰον .v|  
] . αν  
] . . .

The 'Variants' panel at the bottom contains the following fields:

Selected words: ἐφιδεῖν

Apparato papirologico	Annotation	Apparato prosopografico	Annotation
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Lessico medico	Annotation	Apparato varianti	Annotation
ἐφιδεῖν	visitare (effettuare una visita medica)	<input type="text"/>	<input type="text"/>

# Indexes

Image: Referti medici Greci su papiro  
Zoom: 19.7% PSI 3241 [T]

Transcription: Referti medici Greci su papiro  
Column A

1 Πείσιδι [ στρ(ατηγῶ)]  
παρά Θέω[νος τ]οῦ Ἄρπα-  
ήσιος ἰατρο[ῖ]οῦ τῶν ἀπ' Ὀ-  
ξυρυγχῶν πόλεως.  
5 ἐπετ[ρ]άπη μοι ὑπό σου  
διὰ [ . . . . ] ἵς τοῦ Πτ[ι]ολε-  
μαί[ου] ὑπ[η]ρέτου τῆ [ . . .  
ἐνν[ε]ακα[ιδεκάτ]η τ[ι]οῦ]  
ἐ[ν]εστώτος μηνός [Πα]  
10 Σωτηρείου ἐφιδεῖν τὴν  
περὶ . . . . ἰν Ἄρσιήσιος  
διάθεσιν. ἐφιδῶν οὖν  
τῆ αὐτῆ ἡμέρα ἐπακο-  
λουθοῦντος τοῦ ὑπηρε-  
15 τοῦ], εὔρον αὐτὴν ἀπο  
. . . ] . ας νόσου καὶ ἐξαι-  
[μοῦ]σαν. διὸ ἀναφέ-  
[ρω] (ἔτους) ἰε αὐτ[ο]κράτορος  
Ἰούλιου Σεβαστοῦ

Index: T

Open Rebuild Cancel Variants

Index

- T
- PA
- PR
- ME
- V

FigAttVit.doc - Microsoft ...

17.33



# Textual criticism for medieval manuscripts

The screenshot displays a digital manuscript viewer with two main windows. The left window, titled 'Image: Chantilly', shows a page of a medieval manuscript in Gothic script. The text is written in a single column and includes a large initial 'A' for 'Audi parleram'. The right window, titled 'Transcription: Chantilly', shows a modern transcription of the text. A red box highlights the word 'enssens' in the transcription, which corresponds to a word in the manuscript image. Below the transcription window is a 'Linked Documents' panel with a table of links.

Image: Chantilly

Transcription: Chantilly

en bon vin, e d' ayssso beva p(er) lo mati(n):  
lo ventre ti purgara e las humors t' en caussara que cruzas son e del mal ayre. Piza enssens ambe ayga fort e d' ayssso te fay onher fort: ja noyre no(n) ti poyra ni la flayror no(n) soffira. Pren de l' enssens e d' api p(er) egal e beu lo suc p(er) lo gruoch mal: lo fege gueris sen dupta(n)ssa si on lo pren ambe lo nart de Franssa. Al ve(n)tre ajuda fort si on lo pre(n) am vinaygre fort. Si serp mort subitame(n)t ambe lo vin beva d' enssens. Mescla lo suc a(m)be mel cueyt e met en en los huells de nuech car clars los ti fara e l' escurdat t' en caussara. Si las aurellas te clauzen de l' **enssens** las p(er)fuma ben ij jorns o iij ser e matin, cueyt en ayga o en bon vin. Si brach t' en yeis sia pizat l' enssens an mel dessus pauzat. A nafra fresca fa gra(n)t ben enssens pizat sens outra ren. Si on lo met dess(us) a(m)be (un) drap, el tol los flayrorx del cap. Si en l' ayga on cos se banha on gräftigi tol e mermason. Si tu te(m) des sobre mar cant en lo lin volras intrar beu de l' enssens a(m)be vin pur pueys potz anar p(er)

Linked Documents

	E	D
Princeton	E	D
	A	D
	A	D
	A	D
	A	D
	A	D

Link to collated sources

# Analysis of the variant reading in the collated source

lo cor z aune los uerim  
e fai mout pular. cret  
alas femnas las floz.  
A Precobrar manar  
Ad home cant a perdit  
lomaiar pun de l'erba  
blanca so el leixens 7.1.  
puic de cil mouza so el  
erba. e del nardre de  
un aigie z aial tant  
deluna erba con d'iau  
tra epital las efauz las  
fort coue e daquella  
aiga dal h abeure et  
aqui mezeil aua no  
luniat de manar per  
O uy pren purgar uerre  
freny p purgar benne  
preonera peres berz ber  
melhs e adou de peyza  
culy e de penoll. nra d'q  
net ruenas lo cor de d'uo

Ad home d'aureillas.  
que a mal en las aurei-  
llas prin l'eixens e per-  
fuma las aurellas per  
tres ves o per iiij lo  
matin e l ser atressi

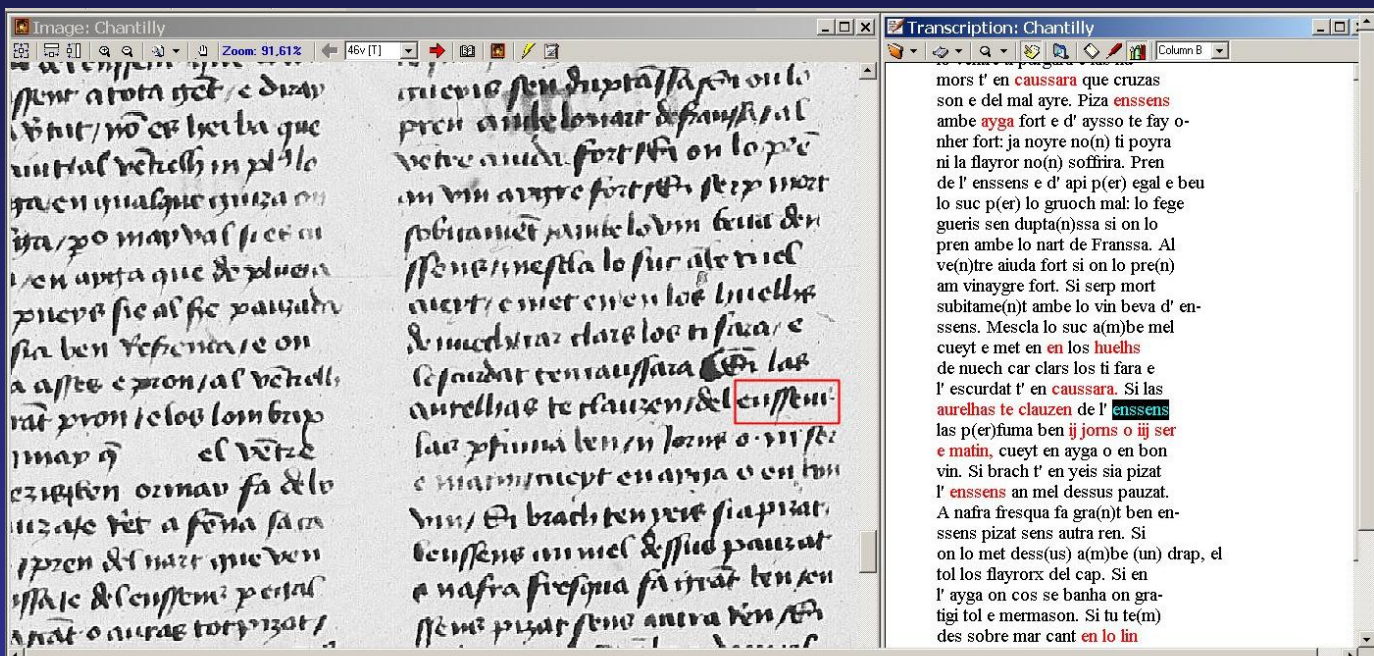
la no i tocara pueissas.  
P(er) mal d'uells.  
Ad home que a mals els  
uuls prin l'eixens e pi-  
sta lo ab mel e tu lo mit  
sus los uuls e gitara fo-  
ras lo mal. Per mal  
Ad home d'aureillas.  
que a mal en las aurei-  
llas prin l'eixens e per-  
fuma las aurellas per  
tres ves o per iiij lo  
matin e l ser atressi  
con l'erba es cremada,  
mais avant vol esser

Bona leyada

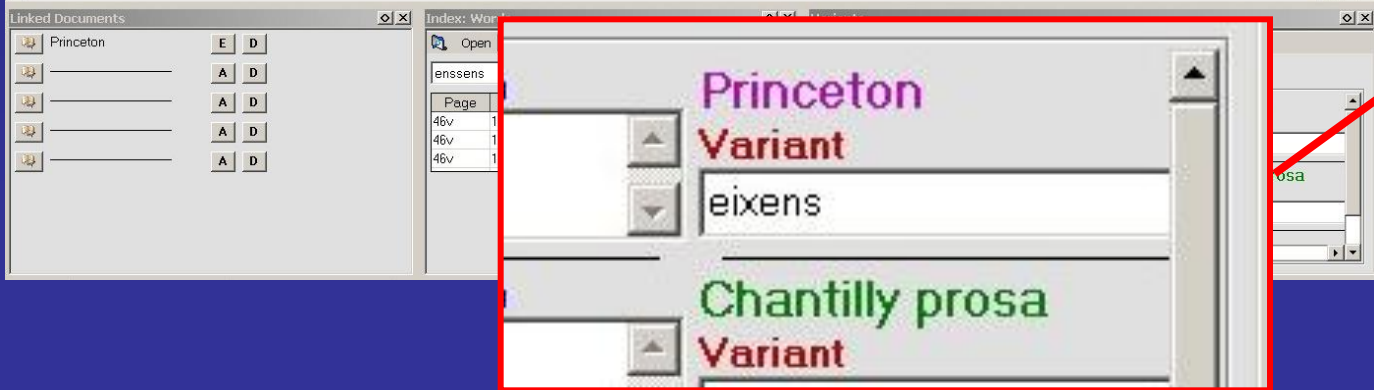
Selection of  
the variant  
*eixens*



# Recording the variant in the apparatus



Memoriz. of  
*Eixens* var.  
in critical  
apparatus



# Variant search in different ancient printed editions of the same work

630 **Contradientium Medicorum,**

alter est tumor exiguus non dico nullus vt Gentilis : & ideo hæc **possunt** quanquam raro cerebro contingere. Sed Erysipelas magis lædit membranas, & fit semper in superficie cerebri, ideo non potest dici apostema cerebri, apostema etiam molle cum diuidat cerebri substantiam, celeriter occidit: ideoque Aëtius non egi de vilo cerebri apostemate, & recte. Ratio Serapionis est bona, tendit enim ad hoc quod non fiat apostema, id est tumor conspicuus & permanens in vna parte, nam partes prociidunt cerebri ob mollitiæ, cum sensu digno tumore inuicē diuelluntur. Auicenna ad contingentiam recurrit, quæ res proculdubio lata est, velut si quis vellet affirmare hominem nasci posse cum iecore in sinistro latere.

**CONTRADIC. XII.**  
*Ventriculus an chilo nutriatur.*

**G**alenus in tertio virtutum naturalium dixit ventriculum sibi primo concoquere, inde sumpto ex eo quod concoxit alimento, reliquum aliis membris impartiri, quam sententiam quoad nutritionis modum declarauit etiam in secunda aphorismorum, tum etiam Auerroës secundo Collectaneorum & Princeps ipse dum dixit occultum autem ventriculi primum nutrimentum

tis quodd omne membrum quod alij præparat totum quod præparat ad illud transmittit: quare si ventriculus præparat cibum & chylum iecori, non retinebit quicquam ex illo, nam vt in exemplo adducit, testes præparant semen glandulæ, ideo nullum habent alium meatum aliud tendentem. Pro prima tamen opinione adducuntur etiam tres alie rationes, prima quia si ex sanguine etiam nutritur, nutritur etiam à chilo & à sanguine, quod esse non potest, ad hoc responderet aliquis quod à solo sanguine nutritur pars interior ventriculi & nullo modo à chilo. Secundo sequeretur quod natura mitteret chylum à ventriculo ad iecur per easdem venas per quas sanguinem traheret. Respondeo quod hoc argumentum potius concludit oppositum, nam si ventriculus nutritur à chilo, idem nutritur & à sanguine & à chilo. Dico igitur quod venæ quæ nutriunt ventriculum, non sunt illæ quæ exugunt chylum, sed chylus transmittitur prius in intestina, & ibi à iecore exugitur per venas mesaraicas, in intestina non in ventriculum terminatas. Tertia ratio est quod vacuum à vacuo traheret, nam cum ventriculus est vacuum traheret à iecore vt nutritur, idem contingeret ipsi iecori. Sed hoc nihil concludit, nam posset trahere vt dixi vtrumque membrum vicissim, sed diuersis modis, ventriculus à iecore per venas portæ fundo illius infertur, & iecur à ventriculo per intestina. Ita quod ab intestini trahat per venas

alter est tumor exiguus non dico nullus ut Gentilis: & ideo hæc **possunt** quanquam raro cerebro contingere. sed erysipelas magis lædit membranas, & fit semper in superficie cerebri, ideo non potest dici apostema cerebri, apostema etiam molle cum diuidat cerebri substantiam, celeriter occidit: ideoque Aëtius non egi de vilo cerebri apostemate, & recte. ratio Serapionis est bona, tendit enim ad hoc quod non fiat apostema, id est tumor conspicuus & permanens in una parte, nam partes prociidunt cerebri ob mollitiem, cum sensu digno tumore inuicem diuelluntur. Avicenna ad contingentiam recurrit, quæ res proculdubio lata est, velut si quis vellet affirmare hominem nasci posse cum iecore in sinistro latere.

contradic. xii.

*ventriculus an chilo nutriatur.*

Galenus in tertio virtutum naturalium dixit ventriculum sibi primo concoquere, inde sumpto ex eo quod concoxit alimento, reliquum aliis membris impartiti, quam sententiam quoad nutritionis modum declarauit etiam in secunda aphorismorum, tum etiam Auerroës secundo collectaneorum & Princeps ipse dum dixit occultum autem ventriculi primum nutrimentum quod est in eo tangit, & per ipsum contactum nutritur, clarus autem in tertiadecima tertii hoc declarauit. sed oppositum habetur tertio de temperamentis cum dicit, oportet, quod nutrit primo in ventriculo & iecore ac venis aliquam traxisse moram: igitur cum chylus non fuerit in iecore, neque in venis nutrire non potest, adducuntur & duae rationes non leues a Iacobo, quarum prima est, quod ex eisdem nutritur dum sumus in ute-

Linked Documents

- Ed. Parisii
- Ed. Marpurgi
- Ed. Lugduni
- Ed. Venetis

Index: Words

Word	1	2	3	4
possunt				
p629	1	4	8	
p629	2	68	3	
p630	1	2	5	
p631	1	18	6	
p631	1	59	2	

Index for...

- Words
- Names
- Variants

Variants

Save Delete Index Levels

Selected words

Annotations

Variant

List of collated editions



# Image of the corresponding page in the selected edition

The image displays a digital edition interface with three main windows:

- Doc: cardano\_venezia - Page: ven.015R**: Shows a scan of a manuscript page. The text is in Latin, discussing the properties of zinziber (ginger) and piper. A red box highlights the phrase "Paulus vero dicit zinziberem obtinere,".
- Transcription: cardano\_pricipale**: Shows the transcription of the manuscript page. The text is in Latin, discussing the properties of zinziber and piper. A red box highlights the phrase "vero dicit, zinziberem obtinere, ut piperem non videatur libere non videatur libere".
- Linked Documents**: Shows a list of linked documents: cardano\_venezia linked, cardano\_lione linked, and cardano\_marpurgi linked.
- Index: word\_latin**: Shows a table with columns for Page, Column, Line, and Position.
- Variants and Annotations**: Shows a table with columns for Personal choice, Variant, Fonte 1, and Fonte 2. The selected word is "obtinere".



## Future activities

- Application to philology of the manuscript text of modern and contemporary authors (critique génétique)
- Possibility of exporting the edited text, variant apparatuses, annotations and indexes (e.g. *index locorum*, *index verborum*) for printed editions
- etc.....

# Pinakes Text and INTEREDITION

**PTK** is a web-based platform with integrated modules for computer-assisted scholarly editing within the *roadmap of electronic publishing*

## **INTEREDITION**

(Interoperable Supranational Infrastructure  
for Digital Editions)

**COST ACTION IS-0704**  
European Science Foundation

# General description of PKT

<http://pinakes.imss.fi.it/index.php/Pinakestext>

## Partners:

- CNR, Istituto di Linguistica Computazionale, Pisa
- Fondazione Rinascimento Digitale, Firenze;
- Istituto e Museo della Storia della Scienza, Firenze;
- Ministero per i Beni Culturali, Roma