Oltre XML?

Problemi di rappresentazione della tradizione fluida

Domenico Fiormonte, Università Roma Tre
(fiormont@uniroma3.it)
Desmond Schmidt, University of Queensland, Australia
(schmidt@itee.uq.edu.au)

Atelier sull'informatica per i testi e i manoscritti medievali
Arezzo, 26-30 novembre 2007

Il testo e i suoi supporti/ le sue rappresentazioni

- “Il fine della filologia è la storia” (Schlegel)
- “la natura del testo è condizionata dai modi della sua produzione e riproduzione, … insomma il testo non è una realtà fisica ma un concetto limite” (Segre).
- L’opera è “una perenne approssimazione al valore” (Contini)
- Le strutture della comunicazione si evolvono e perdono progressivamente il legame con le proprie radici (R. Harris, A. Leroi-Gourhan, ecc.)
- “Ogni limite è un taglio arbitrario entro un insieme continuamente mobile” (Foucault)

Filologia come interfaccia della trasmissione della conoscenza

- Filologia e critica del testo: ricostruzione e interpretazione del testo
- Critica genetica: aspetti visuali e materiali della scrittura
- Filologia digitale/computazionale: codifica come atto ermeneutico, automazione procedure ecdotiche
  - rappresentazione della *mouvance* testuale e del processo di scrittura
- Post-filologia: accesso e fruizione diretta al testo, coautorialità (Wiki, Writeboard, Google Docs)

Testo come archivio

- Archivi testuali marcati / DB
  - BIBIT, BIVIO (Italia)
  - Gallica (Francia)
  - Cervantes (Spagna)
  - Banco Ímagenes Quijote
  - Chaucer, Kleio
- Archivi ipermediali
  - Blake, Rossetti,
- Edizioni collaborative
  - Finnegan, Vergil, Hypernietzsche
- Edizioni/pubblicazioni genetiche / variorum
  - Gide
  - Newton Chymistry
  - Digital Variants
  - Quijote *variorum*

Tecnologie utilizzate

- Il 90% dei prodotti citati usa tecnologie che derivano dai paradigmi testuali o mediali, ovvero:
  - Basi di dati marcate (XML) = organizzazione strutturata dei materiali e possibilità di ricerca, comparazione, analisi
  - Ipertesti e ipermedia = assemblaggio di materiali eterogenei, possibilità di ricerca su elementi non-testuali a partire da metadati
- Pochissimi utilizzano la tecnologia per esplorare le possibilità grafico-visuali o quelle processuali del documento

Il principio metalinguistico

- "La scrittura è metalinguistica" (Olson)
  - Essa è "rappresentazione" e non mera trascrizione del parlato, ne fornisce cioè un modello *concettuale*
  - "…già la trascrizione – ma anche la trasmissione orale – sovrappongono all'idealità del testo una porzione di materializzazione, grafica o fonica, che può provocare degli effetti paratestuali." (Genette)
- In modo simile, il passaggio dall'analogico al digitale descrive un processo di rappresentazione
  - Anche nella codifica di un testo è necessario esprimere un modello concettuale, ovvero astratto, di quel testo
- La codifica digitale è una "rappresentazione" e una "traduzione" di un oggetto in/attraverso un altro linguaggio: un meta-linguaggio

I linguaggi di markup

- Vantaggi dei LM: un sistema di codifica dichiarativo è in grado di rappresentare le caratteristiche di un documento, in modo indipendente da particolari finalità di trattamento, e da contingenti forme di presentazione su un qualsivoglia supporto (stampa, video, sintetizzatore vocale, ecc.)
- Svantaggi: la separazione dell'atto di rappresentazione del documento dalla sua eventuale fruizione è ontologicamente costitutiva dei LM

In conclusione

- Qualsiasi rappresentazione digitale di un testo “sottende una teoria del testo in generale” (Sperberg McQueen)
- I LM costruiscono l’**identità** del documento e insieme il suo **accesso** (es. i metadati)

- Ma siamo sicuri che sia questo il futuro che vogliamo per il documento digitale?

Verso il Web 2.0

- Generatori di testo
  - Insulti shakespeariani, Lettere di protesta,
  - PAC, TagCrowd
- Scritture collettive / Autorialità multipla
  - Newsgroup, MUD, Websapiens
  - Blog, Wikinovela, Literatrónica, LitPart
- Testi processuali
  - FuzzMail, WriteBoard
- Testi visuali
  - What is an author?, Mark America: Filmtext
  - Gutenkarte, AmazType, TextArc, Tierra
- Oltre il testo
  - CaveWriting, Game, game, game
  - La fine del ciberspazio: *Spime*

Web 1.0 +

Web 2.0 +

Digital Variants

- Archivio digitale di artefatti letterari inediti in linea dal 1996
- Studia il processo di scrittura di testi letterari di autori contemporanei
- Conserva immagini e trascrizioni di autografi, avantesti, bozze, ecc.
- Sviluppa soluzioni per la rappresentazione e fruizione delle tradizioni fluide e della genesi testuale

Attuali soluzioni per la digitalizzazione

- Scandire le immagini
  - Pro:
    - Lo scanner cattura l'artefatto così come appare
    - Protegge gli artefatti più fragili e li rende accessibili
  - Contro:
    - Il contenuto non è sempre accessibile
      - Es. il testo non è sempre leggibile
- Codificare il contenuto (XML)
  - Pro:
    - Contenuto accessibile e ricercabile
    - Può essere formattato in modo automatico
  - Contro:
    - La struttura *formale* del markup non sempre riesce a rappresentare la struttura *informale* degli artefatti di scrittura
    - Richiede un alto livello di conoscenze specialistiche agli operatori culturali (filologi, paleografi, storici dell'arte, ecc.)

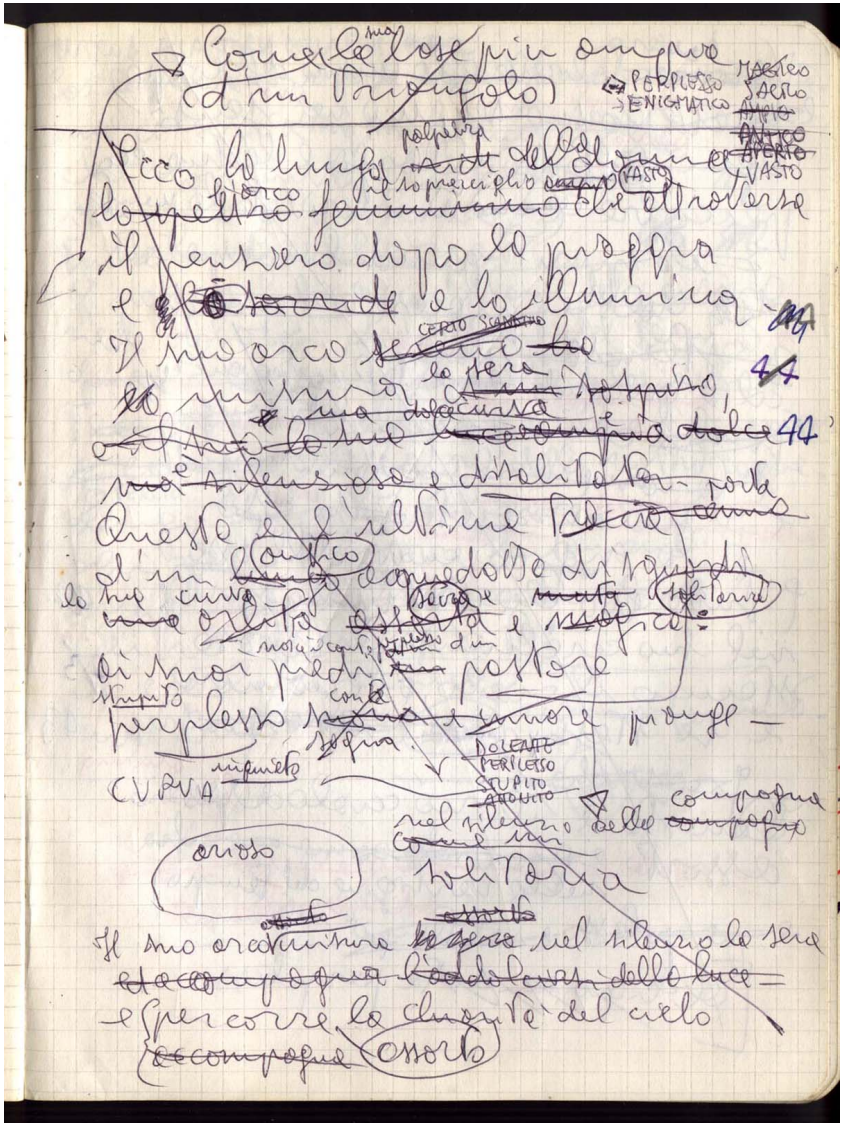
Tutte le soluzioni DV (1996-2007)

- Word:
  - Sanvitale
- HTML
  - Tutti…
- HTML / Java Script
  - Cerami, Sanvitale
- SGML-TEI
  - Sanvitale
- Flip-Zooming
  - Sanvitale
- XML-TEI
  - Magrelli, Cerami
- Flash
  - Magrelli
- MVD
  - Cerami, Magrelli, Galiano

Il markup non è adatto a rappresentare la variazione

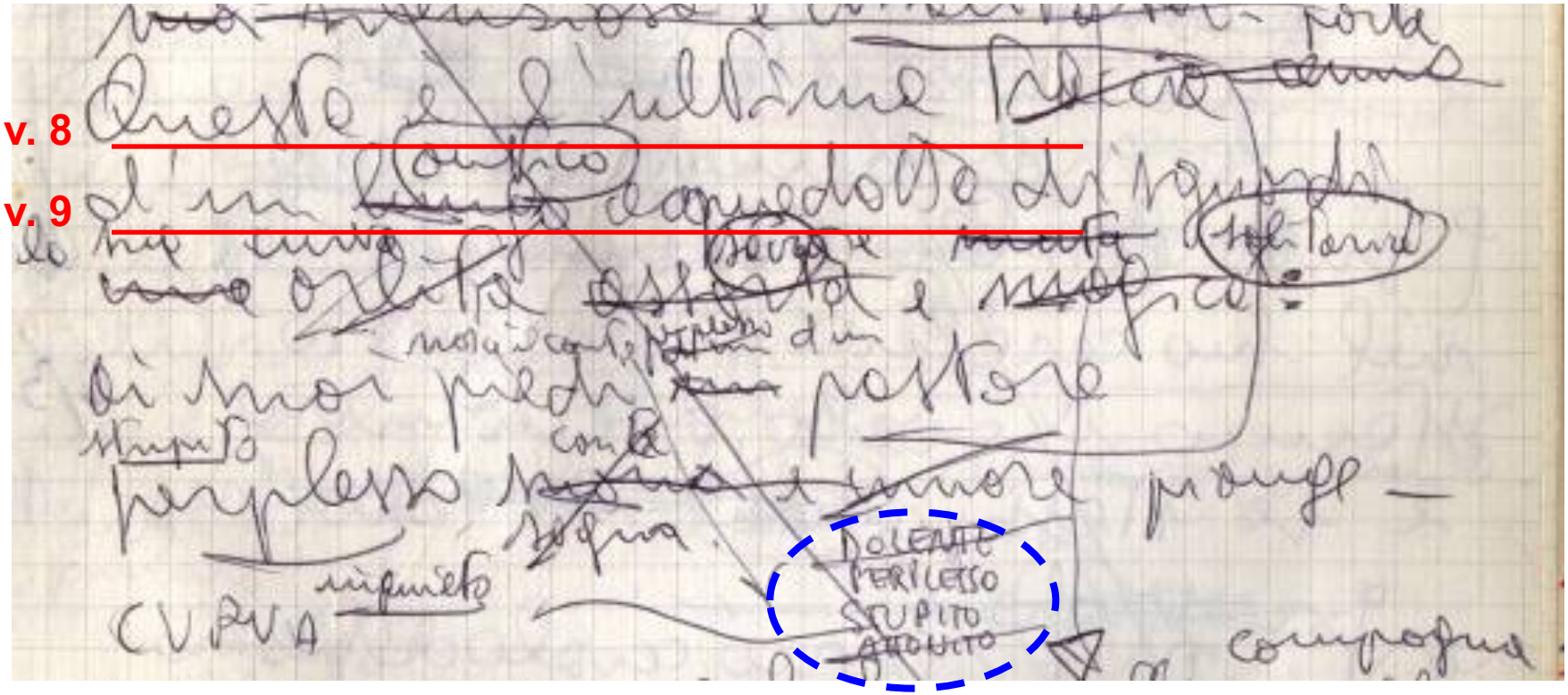
- L'esperienza di DV mostra che la fluidità del processo di scrittura non può essere adeguatamente rappresentata da una struttura gerarchica (Fiormonte e Martiradonna 2007)
- Vetter and McDonald (2003), dopo aver provato ogni metodo per codificare le opere di Emily Dickinson, concludono:
  - 'Ultimately, no entirely satisfactory method of encoding Dickinson's Variants emerges, as more simple coding strategies fail to capture Dickinson's complexities … and more complicated strategies produce awkward and cumbersome code and retain TEI's [i.e. XML's] known difficulties with tagging multiple and overlapping hierarchies.'

Il peggiore dei casi possibili?



Che tipo di modello possiamo sviluppare per rappresentare la struttura testuale di questo oggetto?

V. Magrelli, frammento dell'autografo di *Ecco la lunga palpebra*



Ecco la lunga palpebra della donna,
il sopracciglio vasto che attraversa
il pensiero dopo la pioggia
e lo illumina. Il suo arco
misura nel silenzio la sera
percorrendo assorto
la chiarità curva del cielo.
Questa è l'ultima porta
d'un antico acquedotto di sguardi.
(ver. def. Magrelli 1980)

La ‘vecchia’ idea: XML-TEI

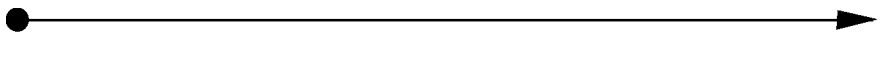
```
<seg type="l">Questa è l'ultima
  <app>
   <rdg varSeq="1">traccia</rdg>
   <rdg varSeq="2"><del hand="M"
    type="overstrike">traccia</del><add hand="M"
    place="right">cenno</add></rdg>
    <rdg varSeq="3"><del hand="M"
    type="overstrike">cenno</del><add hand="M"
    place="supralinear"> porta</add></rdg>
  </app>
</seg><lb/>
<seg type="l"> d'un
  <app>
    <rdg varSeq="1">lungo</rdg>
    <rdg varSeq="2"><del hand="M"
     type="overstrike">lungo</del><add hand="M"
     place="supralinear"><emph rend="circle"
     id="adj_2">antico</emph></add></rdg>
   </app> acquedotto di sguardi,
</seg><lb/>
```

```
<seg type="l">ai suoi piedi
 <app>
   <rdg varSeq="1">un pastore</rdg>
   <rdg varSeq="2"><add hand="M" place="supralinear">nasce il canto
     <note type="arrow" resp="M" place="foot">
      <list type="simple" id="adj_list2">
        <item><del hand="M" type="overstrike">DOLENTE</del></item>
        <item>PERPLESSO</item>
        <item>STUPITO</item>
        <item><del hand="M" type="overstrike">ATTONITO</del></item>
      </list>
     </note>
   d'un pastore</add></rdg>
  <rdg varSeq="3">nasce il canto
     <note type="arrow" resp="M" place="foot">
      <list type="simple" id="adj_list2">
        <item><del hand="M" type="overstrike">DOLENTE</del></item>
        <item>PERPLESSO</item>
        <item>STUPITO</item>
        <item><del hand="M" type="overstrike">ATTONITO</del></item>
      </list>
     </note>
    <del hand="M" type="overstrike">d'un</del>
    <add hand="M" place="supralinear">perplesso d'un</add>pastore</rdg>
 </app>
```

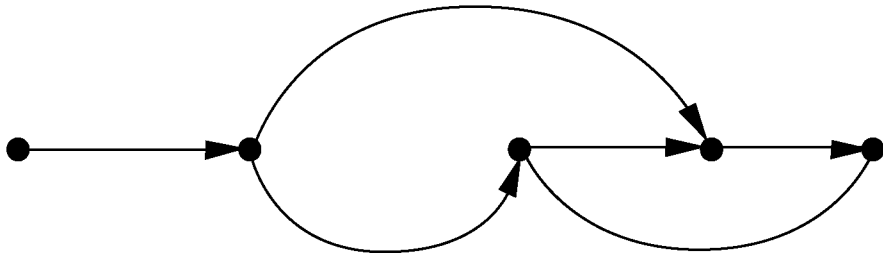
Un nuovo modello: il testo come rete

- Partendo da una concezione diversa di testo è possibile arrivare a un diverso modello di rappresentazione digitale
- Il markup, nonostante la complessità crescente, rimane un testo lineare
- E se invece il testo si diramasse e convergesse (*merge*) in risposta alla varianza testuale?
- Allora sarebbe possibile rappresentare diverse versioni o 'viste' dello stesso documento in un unico file

Dunque da questo:



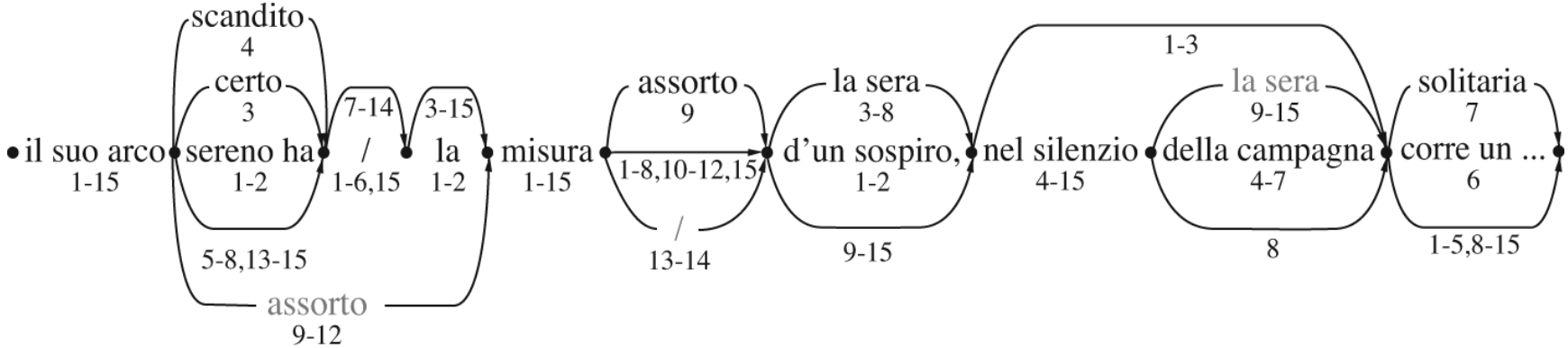
Passiamo a questo:



Il *grafo variante*

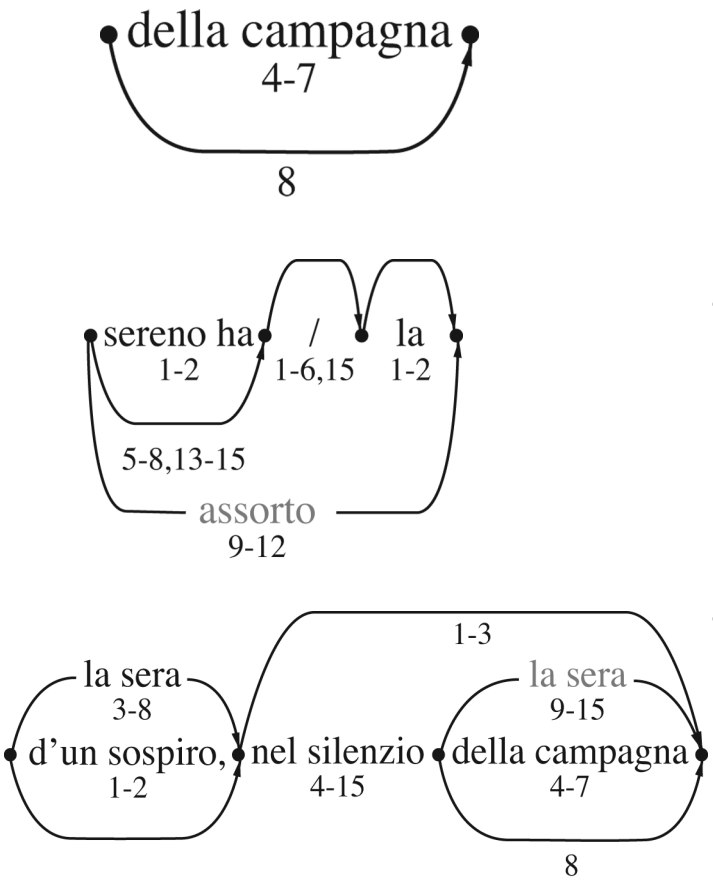
1. Autografo originale (3 artefatti, in tutto 15 versioni/livelli)

2. Rappresentazione a grafo o a rete



(I numeri rappresentano la versione)

In questa rappresentazione abbiamo…



- Inserimenti o cancellature sono rappresentate da archi vuoti. L'arco etichettato '8' è vuoto, cioè in quella versione non c'è un testo che corrisponda a 'della campagna' delle versioni 4-7.

- Le varianti sono percorsi alternativi fra due medesimi punti. La parola 'assorto' delle versioni 9-12, è una variante di 'sereno ha/la' delle versioni 1-2.

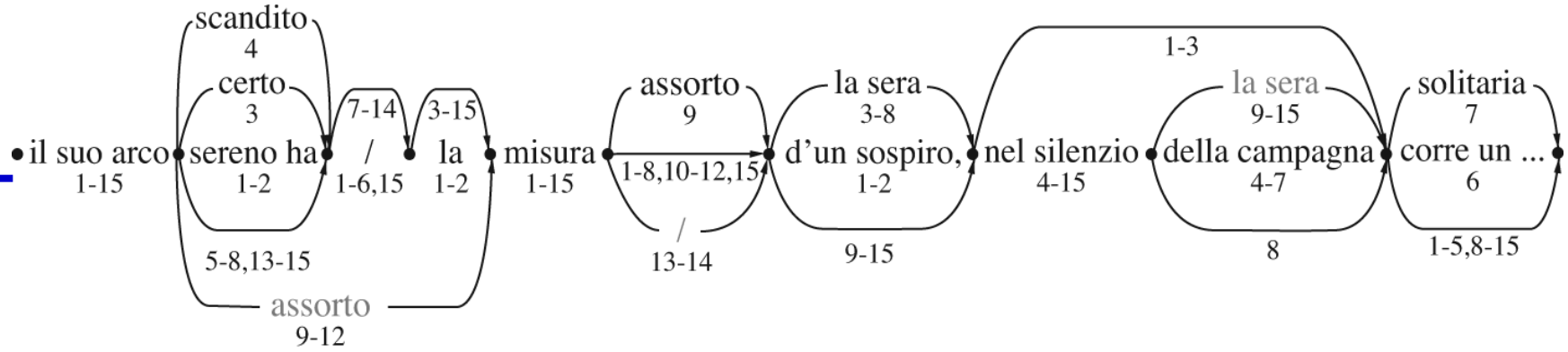
- Le trasposizioni sono coppie di inserimenti e cancellature dove il testo inserito o cancellato è lo stesso. Nelle versioni 9-15 'la sera' è trasposto da prima di 'nel silenzio' al nodo successivo.

Rappresentazione: trascrivere gli archi del grafo in un ordine specifico

- Chiamiamo tale ordine *Pairs-list representation*. Pensiamo a una pila di monete: ciascuna di esse ha testa e croce.
- Un Grafo Variante rappresentato in questo modo è una pila o lista di coppie, dove a ciascuna coppia è associato un elemento di testo (che può essere vuoto) e un set di versioni a cui esso appartiene.
- Questo è tutto ciò di cui abbiamo bisogno per rappresentare un grafo variante, giacché la struttura è interamente compresa (*implied*) nella sequenza delle coppie e nei set di versioni.

1-15 Il suo arco
1-2 sereno ha
3 certo
4 scandito
5-8,13-15
9-12 assorto
1-6,15 /
7-14
1-2 la
3-15
1-15 misura
1-13
9 assorto

Il grafo è una struttura bidimensionale!



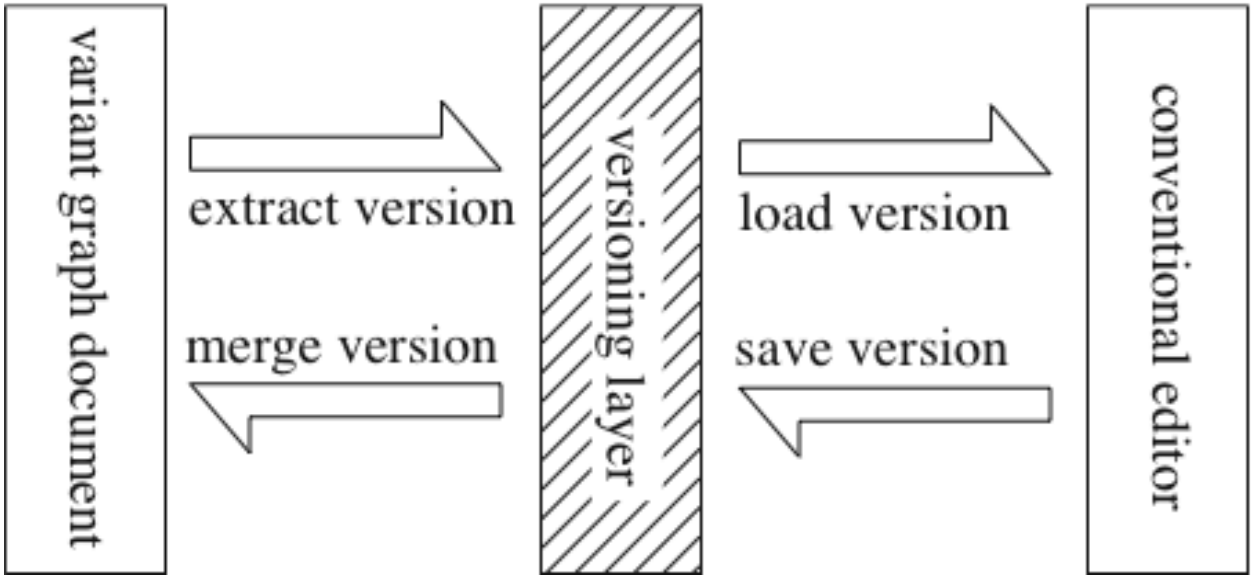
1-15 Il suo arco
1-2 sereno ha
3 certo
4 scandito
5-8,13-15
9-12 assorto
1-6,15 /
7-14
1-2 la
3-15
1-15 misura
1-13
9 assorto

Ricapitolando: vantaggi della *Pairs-list Representation*

- Semplice: una lista di elementi di testo e il set di versioni al quale appartiene
- Può memorizzare migliaia di versioni senza rischio di *overload*
- Può rappresentare multiple versioni *o* multiple codifiche dello stesso documento/artefatto
- Costruisce un singolo oggetto digitale per ciascun documento/artefatto
- Inoltre…

Separa variazione e contenuto

- Un grafo variante separa chiaramente il contenuto di un documento dalla sua struttura sovrapposta (overlapping): le etichette del grafo sono il suo 'testo' e la struttura la sua variazione
- Questo permette di adottare qualsiasi tecnologia (formato binario, XML, PDF etc.) per rappresentare il contenuto, e utilizzare strumenti convenzionali per editarlo



MVD: Un’opera = un documento

- Dunque abbiamo incapsulato tutto in un unico documento, che chiamiamo documento multi-versione (MVD)
- Per ora è disponibile in due gusti: binario e XML (solo per testi). Il formato binario funziona per qualsiasi oggetto ed è assai più leggero

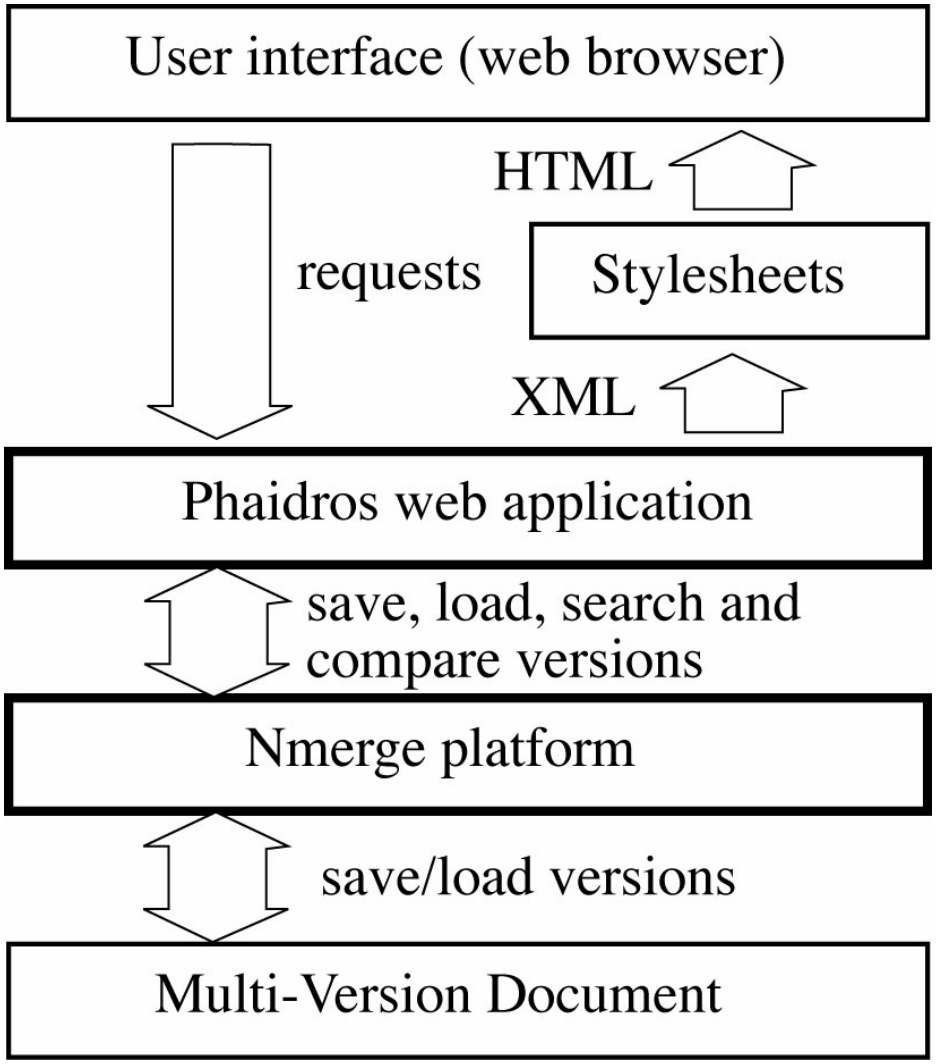
Operazioni possibili su un MVD

- **Ricerca:** usiamo una tecnica di ricerca classica, dove però il meccanismo di ricerca segue le diramazioni e i ricongiungimenti ('branch and rejoin') del testo - in pratica il meccanismo è quasi efficiente come una ricerca lineare
- **Comparare** qualsiasi versione A con una versione B
- **Estrarre** una singola versione
- **Aggiungere** una nuova versione al grafo o editarlo: può essere dispendioso, ma è possibile farlo seguendo alcune tecniche usate in biologia. Nella maggioranza dei casi comunque il grafo verrà letto, non costruito. Dunque si tratta di un costo che possiamo trascurare.

MVD può essere una soluzione per tutti i casi di variazione testuale

- Il formato MVD possiede due caratteristiche che lo rendono adatto per es. alla complessa tradizione dei testi classici:
  - È possibile gestire le tradizioni frammentarie, es.
    - Collazione di manoscritti perduti
    - Papiri
    - Citazioni in scholia o autori antichi
    - Apparato critico di un'altra edizione
    - Congetture isolate
    - ecc.
  - Le versioni possono essere raggruppate gerarchicamente, es. tipo di manoscritto (antiquores, recentiores), papiri, fonti secondarie, altre edizioni, ecc.

MVD sul Web



- È già in funzione un prototipo di web service
- Sarà un ambiente interattivo dove gli utenti potranno **direttamente** manipolare e gestire i documenti multi-versione

Demo

Work: Sibylline Gospel

Recension A
✓ A
E
Me
Vo
Recension B
C1
C2
D
L1
L2
L3
L4
L5
L6
L8
V
Recension C
C3
C4
C5
C6
C7
C8
L9
L10
L11
L12
N
M

Angers, Bibl. Publ. 26 (22), saec. XIII

t Caesari Augusto celebre nomen et regnabit in
t omnem terram sibi. Post haec convenient
aeorum contra Iesum quod multa signa faciet et
eum.

i non vel pauci, non tres neque duo sunt dii,
deus solus, immensus, qui fecit caelum,
ellas et lunam frugiferamque terram et
mare.
n illum honorant aeternam vitam
unt, perpetuum cuius ipsi hereditabunt
n simul amoenissimum hortum.
m vero resurrectio erit et claudorum cursus
us erit; surdi audient et caeci videbunt,
muti.
s quinque simul et piscibus marinis duobus
ia hominum saturabit et colligent reliquias
orum et duodecim cophinos exinde replebunt.
opulorum veritatem docebit
reque furens sedabit; pedibus pacis calcans
que super undas; infirmitatem hominum
rgere faciet mortuos, repellens multos dolores
e vivo saturatio erit vivorum.
haec omnia fuerint perfecta quae dixi, in ipso
olvetur lex.
Et in manus impias infidelium postea veniet.
Dabunt deo alapas manibus incestis et impuro ore
spuent venenatos sputos. Dabit vero ad verbera
sanctum dorsum.

Search: > >>

C1 Cambridge, Univ. Lib. Mm V.29, saec. XII^{med}

In diebus illis erit Caesari Augusto celebre nomen et regnabit in Roma et subiciet omnem **terra** sibi. Post haec convenient sacerdotes Hebraeorum contra Iesum quod multa signa faciet et comprehendent eum.

Non multi non vel pauci, non tres neque duo sunt dii, sed unus solus **deus**, **inmensus**, qui fecit caelum, solem, stellas et lunam frugiferamque terram et undosum mare.
Qui autem illum honorant aeternam vitam hereditabunt, perpetuum cuius ipsi hereditabunt paradisum simul amoenissimum hortum.
Mortuorum vero resurrectio erit et claudorum cursus velocissimus erit; surdi audient et caeci videbunt, loquentur **non loquentes**.
Et **dapibus** quinque simul et piscibus marinis hominun **quinque milia** saturabit**;** et colligent reliquias **fragminum** duodecim cophinos exinde replebunt **in spe** populorum**.**
Ventos compescens verbo mareque furens sedabit; pedibus**que** pacis calcans
ambulansque super undas; infirmitatem **hominibus** solvet, surgere faciet mortuos, repellens multos dolores et de pane vivo **erit** saturatio vivorum.
Sed cum omnia **haec** fuerint perfecta quae dixi, in ipso omnis resolvetur lex.
Et in manus impias infidelium postea veniet.
Dabunt deo alapas manibus incestis et impuro ore spuent venenatos sputos. Dabit vero ad verbera

Search: > >>