Automated Collation

Interedition

Tara L. Andrews
tara.andrews@kps.unibe.ch

Leif-Jöran Olsson
leif-joran.olsson@svenska.gu.se

David J. Birnbaum
djbpitt@gmail.com

Joris J. van Zundert
joris.van.zundert@huygens.knaw.nl

Ronald Haentjens Dekker
ronald.dekker@huygens.knaw.nl

Automated Collation

Interedition

Tara L. Andrews
tara.andrews@kps.unibe.ch

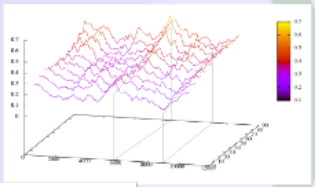
Leif-Jöran Olsson
leif-joran.olsson@svenska.gu.se

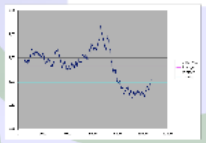
David J. Birnbaum
djbpitt@gmail.com

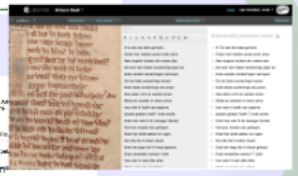
Joris J. van Zundert
joris.van.zundert@huygens.knaw.nl

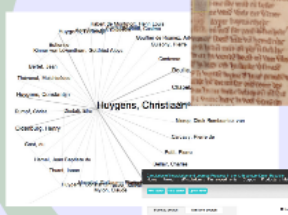
Ronald Haentjens Dekker
ronald.dekker@huygens.knaw.nl

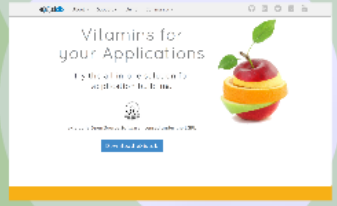
Who are we?



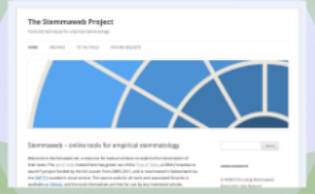


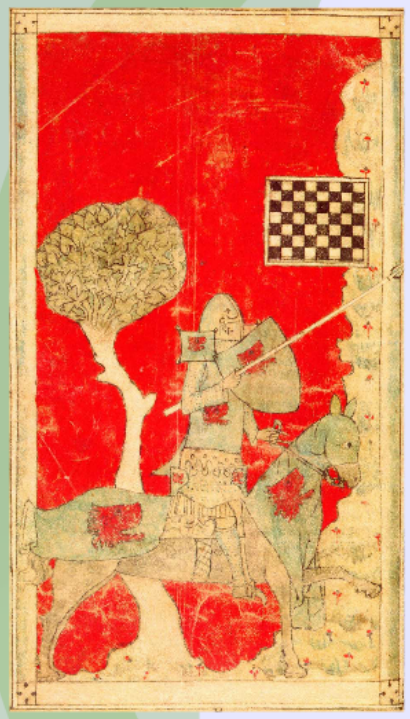


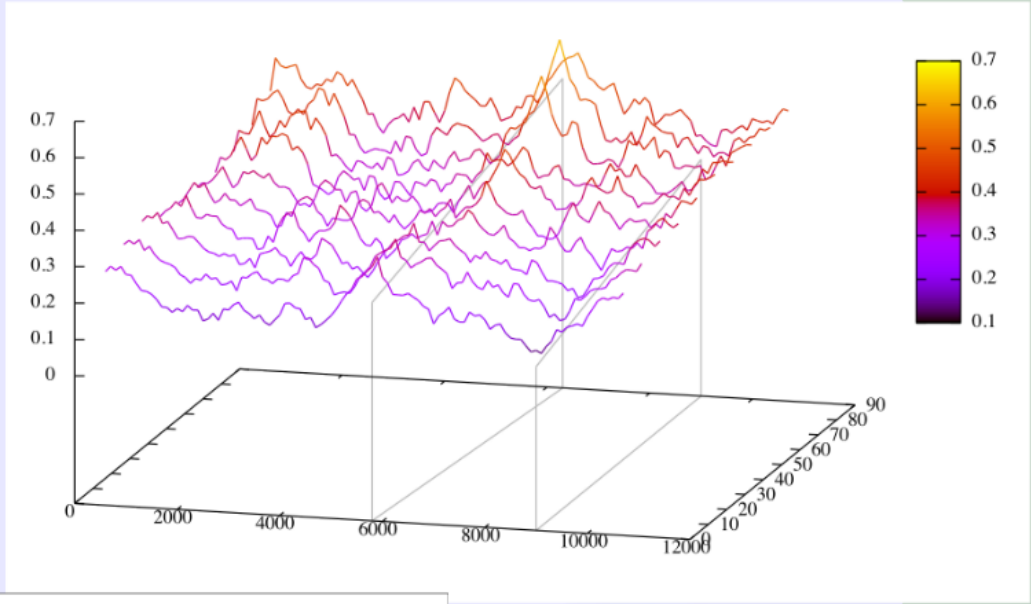

Huygens, Christiaan

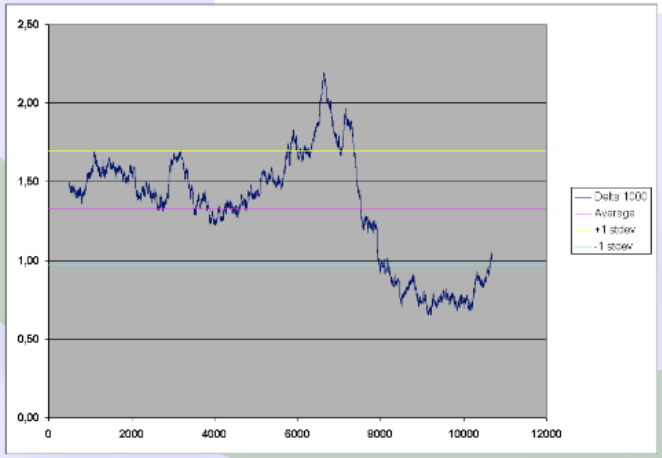

Vitamins for
your Applications

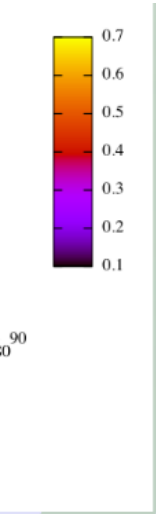
Daniel the Prisoner: A virtual florilegium

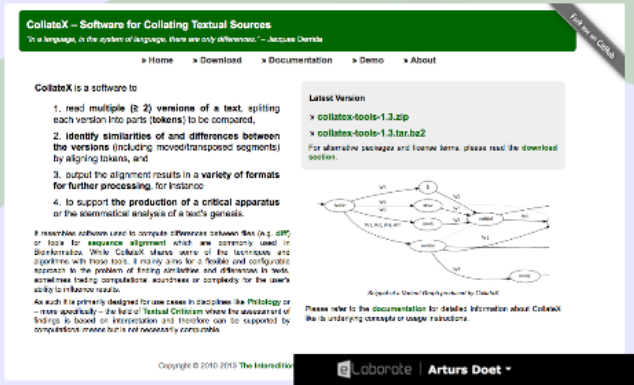

The Stemmaweb Project

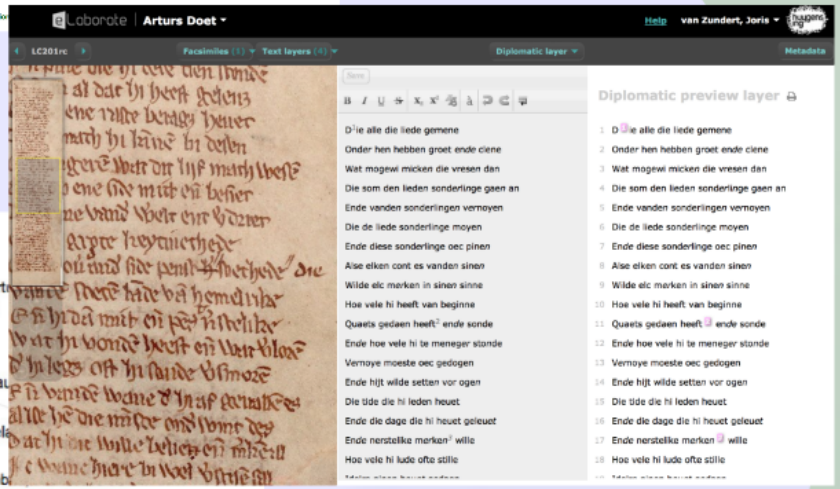


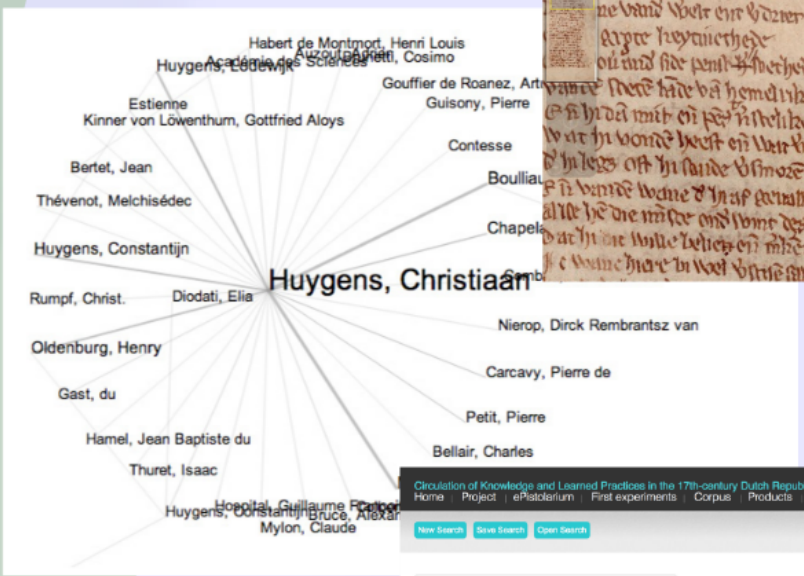

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2000
4000
6000
8000
10000
12000
90
80
70
60
50
40
30
20
10
0

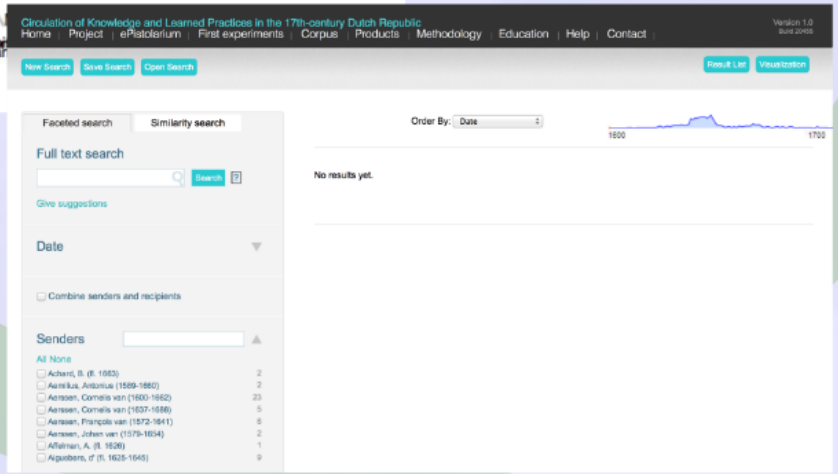

2,50
2,00
1,50
1,00
0,50
0,00
0
2000
4000
6000
8000
10000
12000
Delta 1000
Average
+1 stdev
-1 stdev

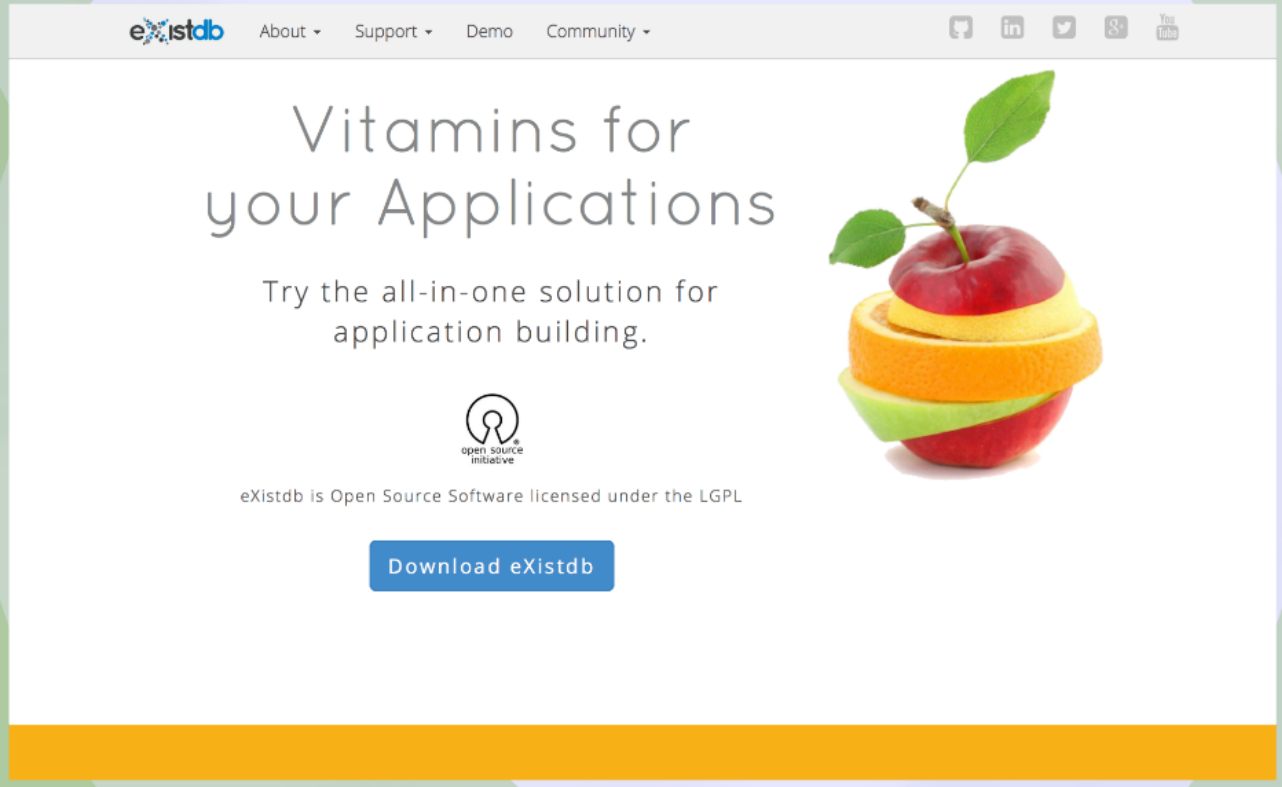

0.7
0.6
0.5
0.4
0.3
0.2
0.1

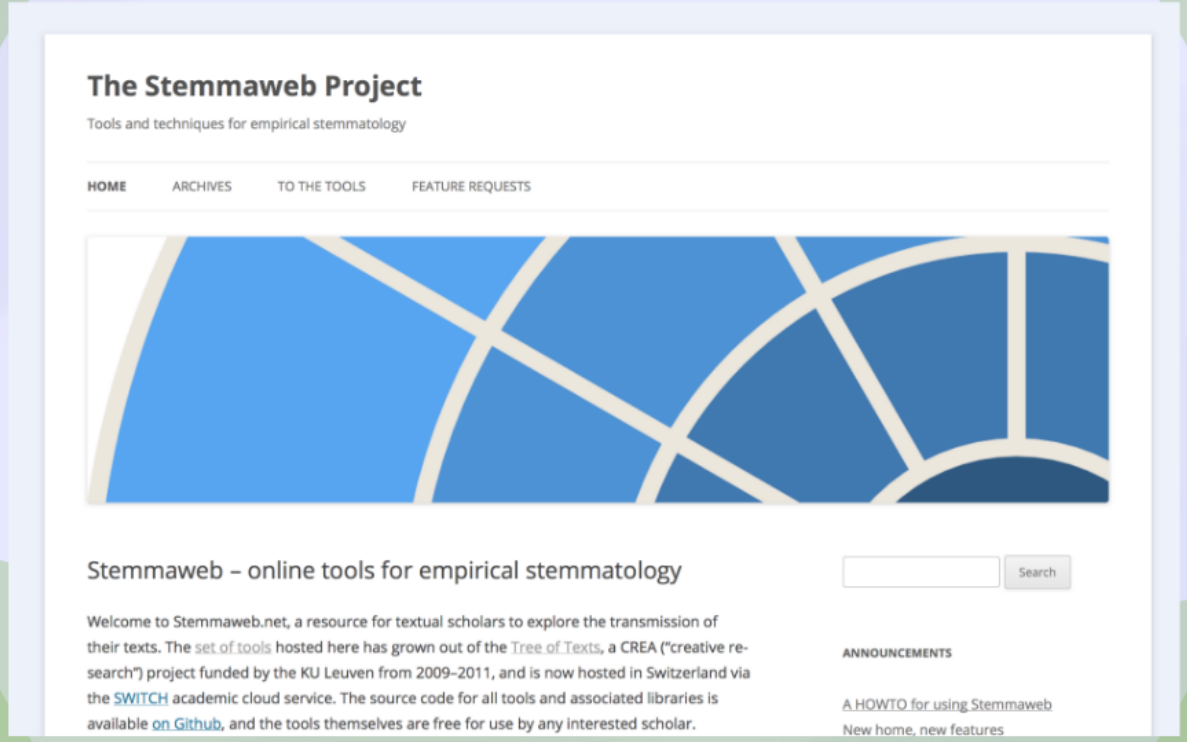

CollateX – Software for Collating Textual Sources
» Home
» Download
» Documentation
» Demo
» About
CollateX is a software to
Latest Version
» collatex-tools-1.3.zip
» collatex-tools-1.3.tar.bz2


eLaborate
Arturs Doet
Help
van Zundert, Joris
Diplomatic preview layer
Die alle die liede gemene
Onder hen hebben groet ende clene
Wat mogewi micken die vresen dan
Die som den lieden sonderlinge gaen an
Ende vanden sonderlingen vernoyen
Die de liede sonderlinge moyen
Ende diese sonderlinge oec pinen
Alse elken cont es vanden sinen
Wilde elc merken in sinen sinne
Hoe vele hi heeft van beginne
Quaets gedaen heeft ende sonde
Ende hoe vele hi te meneger stonde
Vernoye moeste oec gedogen
Ende hijt wilde setten vor ogen
Die tide die hi leden heuet
Ende die dage die hi heuet geleuet
Ende nerstelike merken wille
Hoe vele hi lude ofte stille


Habert de Montmort, Henri Louis
Huygens, Lodewijk
Gouffier de Roanez, Art
Guisony, Pierre
Estienne
Kinner von Löwenthum, Gottfried Aloys
Contesse
Bertet, Jean
Thévenot, Melchisédec
Huygens, Constantijn
Huygens, Christiaan
Rumpf, Christ.
Diodati, Elia
Nierop, Dirck Rembrantsz van
Oldenburg, Henry
Carcavy, Pierre de
Gast, du
Petit, Pierre
Hamel, Jean Baptiste du
Bellair, Charles
Thuret, Isaac
Mylon, Claude


Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic
Home
Project
ePistolarium
First experiments
Corpus
Products
Methodology
Education
Help
Contact
New Search
Save Search
Open Search
Result List
Visualization
Faceted search
Similarity search
Full text search
Search
Give suggestions
Date
Combine senders and recipients
Senders
Order By: Date
No results yet.


eXistdb
About
Support
Demo
Community
Vitamins for your Applications
Try the all-in-one solution for application building.
open source initiative
eXistdb is Open Source Software licensed under the LGPL
Download eXistdb



The Stemmaweb Project

Tools and techniques for empirical stemmatology

HOME ARCHIVES TO THE TOOLS FEATURE REQUESTS

Stemmaweb – online tools for empirical stemmatology

Search

Welcome to Stemmaweb.net, a resource for textual scholars to explore the transmission of their texts. The set of tools hosted here has grown out of the Tree of Texts, a CREA ("creative re-search") project funded by the KU Leuven from 2009–2011, and is now hosted in Switzerland via the SWITCH academic cloud service. The source code for all tools and associated libraries is available on Github, and the tools themselves are free for use by any interested scholar.

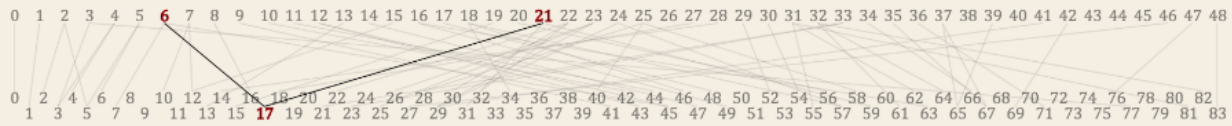
ANNOUNCEMENTS

A HOWTO for using Stemmaweb

New home, new features

<oo>→<дз> Daniel the Prisoner: A virtual florilegium

About



Slovo

Chapter 1

(5) Lo, I wrote, fleeing from the face of my craft, like Hagar the slave-girl from Sarah her mistress.

(6) But I saw, O lord, your good-heartedness toward me and I ran to your customary love. For Scripture speaks: give to him who asks of you, open to him who knocks, lest you be deprived of the Kingdom of Heaven. For it is written: cast your sorrow on the Lord, and He will sustain you in the ages.

Chapter 2

(7) For I am, O prince, O lord, like pale grass, growing by a wall, upon which neither the sun shines nor the rain falls; thus I too am offended by all, because I am fenced-in by fright of your threat, as by a firm bulwark.

[Sir 4:35] [Mt 7:7] [Lk 11:9] [Mt 7:21] [Ps 54:23]

Molenie

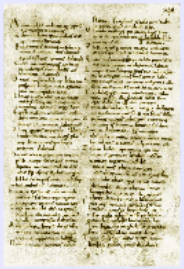
Chapter 6

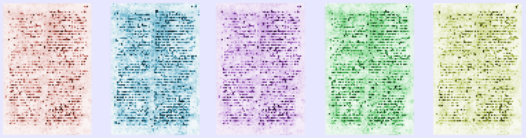
(16) O my prince, O lord! Do not look upon me like a wolf at a lamb, but look upon me like a mother at an infant. Look upon the birds of Heaven, as they neither sow, nor reap, nor gather in granaries, but hope on God's mercy.

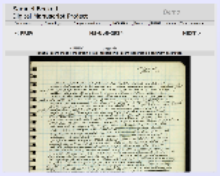
(17) Let not your hand be closed from giving to the poor; for it is written: give to him who asks of you, open to him who knocks, that you not be deprived of the Kingdom of Heaven; for it is written: cast your sorrow on the Lord, and He will sustain you in the ages.

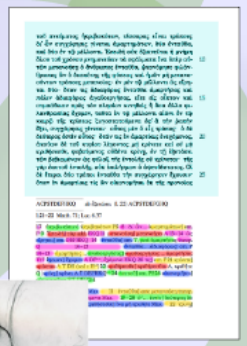
(18) Do not deprive the destitute wise man of bread, nor raise up the thoughtless wealthy man to the clouds. For the destitute but wise man is like gold in a dirty vessel; but the thoughtless wealthy man is like a brocaded pillow stuffed with straw.

The Problem

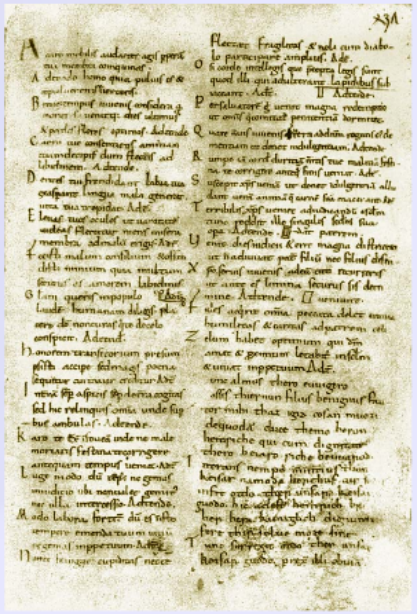


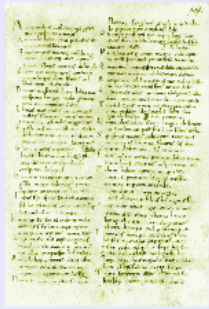


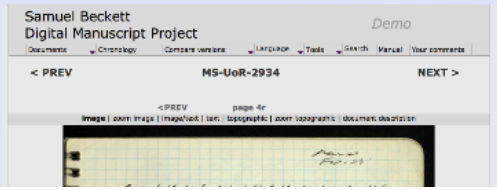


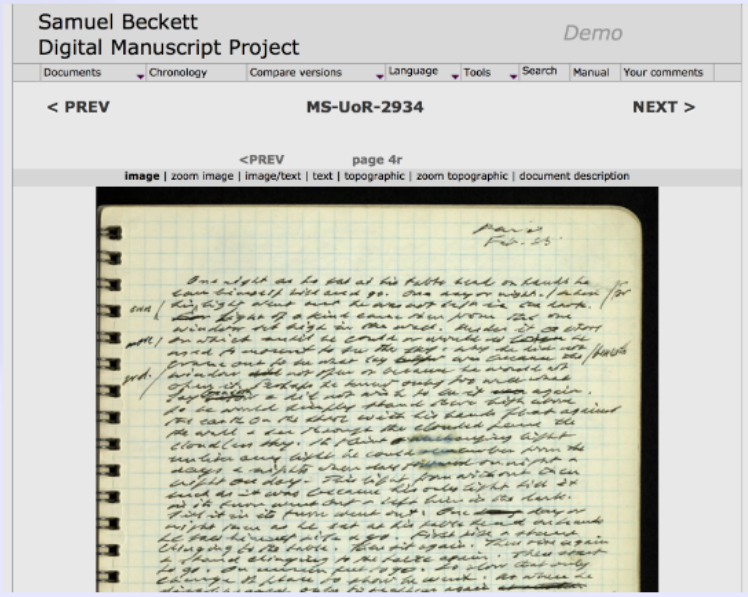


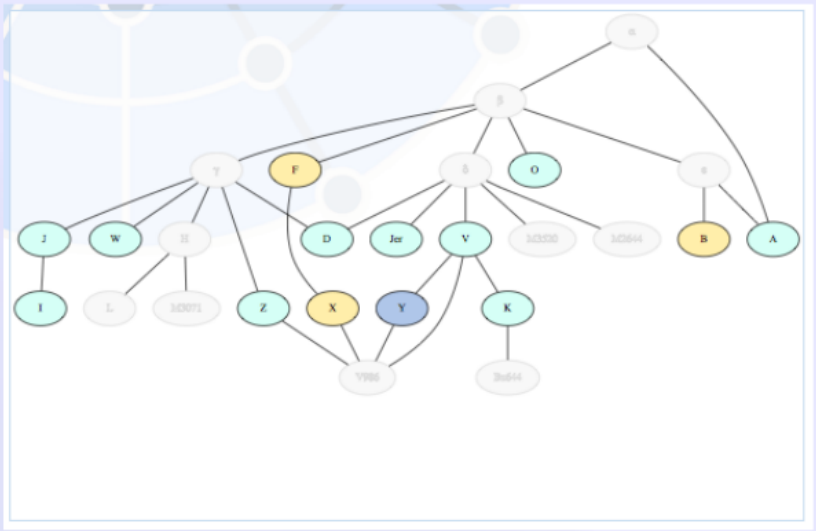
The Problem






Samuel Beckett
Digital Manuscript Project
Demo
< PREV
MS-UoR-2934
NEXT >


Samuel Beckett
Digital Manuscript Project
Demo
Documents
Chronology
Compare versions
Language
Tools
Search
Manual
Your comments
< PREV
MS-UoR-2934
NEXT >
<PREV
page 4r
image | zoom image | image/text | text | topographic | zoom topographic | document description


F
O
J
W
D
V
B
A
I
Z
X
Y
K

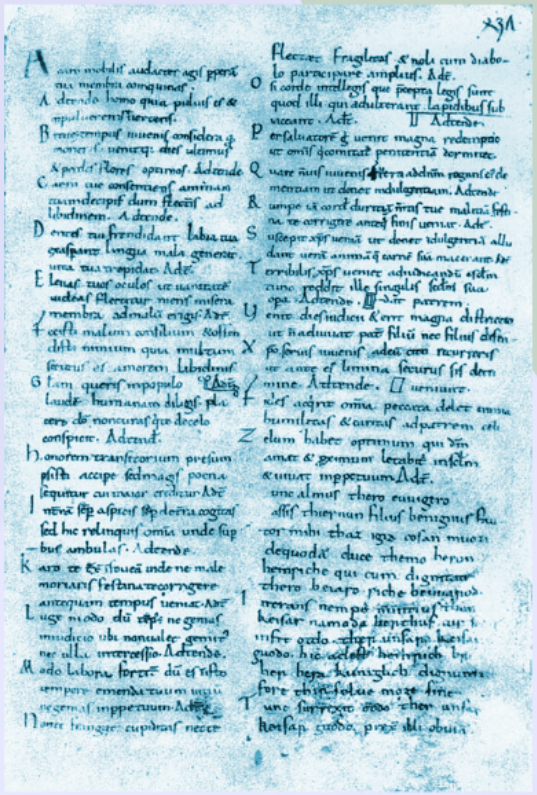
τοῦ πνεύματος ἠκριβωκότων, τέσσαρας εἶναι τρόπους δι' ὧν συγχώρησις γίνεται ἁμαρτημάτων, δύο ἐνταῦθα, καὶ δύο ἐν τῷ μέλλοντι. Ἐπειδὴ οὐκ ἐξικνεῖται ἡ μνήμη ὅλου τοῦ χρόνου μνημονεύειν τὰ σφάλματα ἵνα ὑπὲρ αὐτῶν μετανοήσῃ ὁ ἄνθρωπος ἐνταῦθα, ᾠκονόμησε φιλάνθρωπος ὢν ὁ δεσπότης τῆς φύσεως καὶ ἡμῶν μὴ μετανοούντων τρόπους μετανοίας· ἐν μὲν τῷ μέλλοντι ὡς εἴρηται δύο· ὅταν τις ἀδιαφόρως ἐνταῦθα ἁμαρτήσας καὶ πάλιν ἀδιαφόρως ἀγαθοεργήσας, εἴτε εἰς οἶκτον καὶ συμπάθειαν πρὸς τὸν πλησίον κινηθεὶς ἢ ὅσα ἄλλα φιλανθρωπίας ἔχομεν, ταῦτα ἐν τῷ μέλλοντι αἰῶνι ἐν τῷ καιρῷ τῆς κρίσεως ζυγοστατούμενα ἐφ' ᾧ τὴν ῥοπὴν ἕξει, συγχώρησις γίνεται· οὗτος μὲν ὁ εἷς τρόπος· ὁ δὲ δεύτερος ἐστὶν οὗτος· ὅτάν τις ἐν ἁμαρτίαις ἐνεχόμενος, ἀκούων δὲ τοῦ κυρίου λέγοντος *μὴ κρίνετε καὶ οὐ μὴ κριθήσεσθε*, φοβούμενος οὐδένα κρίνῃ, ἐν τῇ ἐξετάσει τῶν βεβιωμένων ὡς φύλαξ τῆς ἐντολῆς οὐ κρίνεται· τῆς γὰρ ἑαυτοῦ ἐντολῆς, οὐκ ἐπιλήσμων ὁ ἀψευδέστατος. Οἱ δὲ ἕτεροι δύο τρόποι ἐνταῦθα τὴν συγχώρησιν ἔχουσιν· ὅταν ἐν ἁμαρτίαις τίς ὢν οἰκονομῆται ἐκ τῆς προνοίας

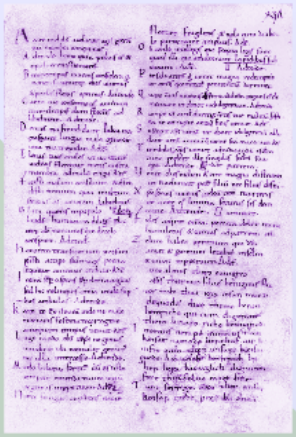
ACPSTDEFHKQ ab ἐξετάσει (l. 22) ACPSTDEFGHQ

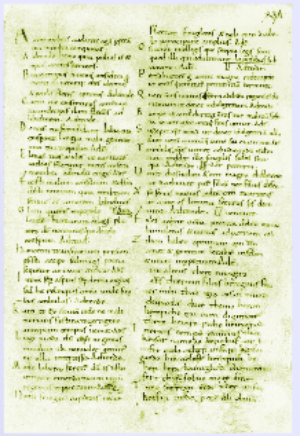
1.21–22 Matth. 7.1; Luc. 6.37

1.7 ἠκριβωκότων] ἠκριβηκότων PS 8 δι' ὧν…ἁμαρτημάτων] om. P 9 Ἐπειδὴ] γὰρ add. EKQ 11 μετανοήσῃ] μετανοήσει A 13–14 ὡς εἴρηται] om. DEFHKQ 14 ἐνταῦθα] om. T, post ἁμαρτήσας transp. Q 14–15 ἐνταῦθα…ἀδιαφόρως] om. P 14–15 ἁμαρτήσας…ἀγαθοεργήσας] ἀγαθοεργήσας … ἁμαρτήσας T 17 ἔχομεν] ἔχωμεν A DF$^{2.c.}$, ἐχόμενα EKQ 20 τις] om. F 21 κρίνετε] κρίνεται A T DH (sed ε H$^{s.l.}$) 22 κριθήσεσθε] κριθήσεσθαι A, κριθῆτε Q | κρίνῃ] κρίνει A T DEFHKQ$^{a.c.}$ 24 ἑαυτοῦ] om. PS 26 οἰκονομῆται] οἰκονομεῖται S T DEFGH

1.9 Ἐπειδὴ] γὰρ add. Max. 11 ἐνταῦθα] ante μετανοήσῃ transp. Max. 17 ἔχομεν] ἐχόμενα Max. 19–20 ὁ^{2}… ἐστὶν] δεύτερος δὲ Max. 21–22 καὶ… κριθήσεσθε] ἵνα μὴ κριθῆτε Max. 22 κρίνῃ] κρίνει Max.







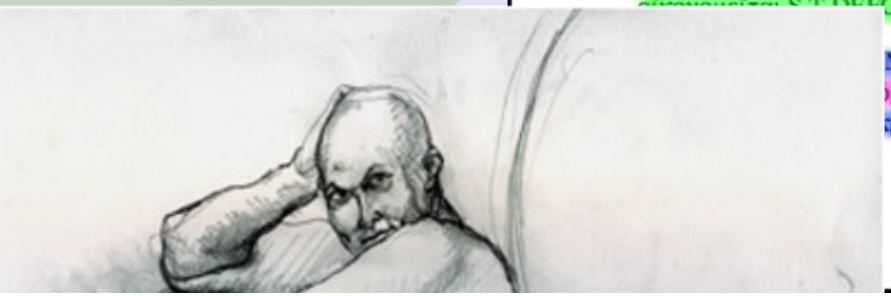
τοῦ πνεύματος ἠκριβωκότων, τέσσαρας εἶναι τρόπους δι' ὧν συγχώρησις γίνεται ἁμαρτημάτων, δύο ἐνταῦθα, καὶ δύο ἐν τῷ μέλλοντι. Ἐπειδὴ οὐκ ἐξικνεῖται ἡ μνήμη ὅλου τοῦ χρόνου μνημονεύειν τὰ σφάλματα ἵνα ὑπὲρ αὐτῶν μετανοήσῃ ὁ ἄνθρωπος ἐνταῦθα, ᾠκονόμησε φιλάνθρωπος ὢν ὁ δεσπότης τῆς φύσεως καὶ ἡμῶν μὴ μετανοούντων τρόπους μετανοίας· ἐν μὲν τῷ μέλλοντι ὡς εἴρηται δύο· ὅταν τις ἀδιαφόρως ἐνταῦθα ἁμαρτήσας καὶ πάλιν ἀδιαφόρως ἀγαθοεργήσας, εἴτε εἰς οἶκτον καὶ συμπάθειαν πρὸς τὸν πλησίον κινηθεὶς ἢ ὅσα ἄλλα φιλανθρωπίας ἔχομεν, ταῦτα ἐν τῷ μέλλοντι αἰῶνι ἐν τῷ καιρῷ τῆς κρίσεως ζυγοστατούμενα ἐφ' ἃ τὴν ῥοπὴν ἕξει, συγχώρησις γίνεται· οὗτος μὲν ὁ εἷς τρόπος· ὁ δὲ δεύτερος ἐστὶν οὗτος· ὅτάν τις ἐν ἁμαρτίαις ἐνεχόμενος, ἀκούων δὲ τοῦ κυρίου λέγοντος *μὴ κρίνετε καὶ οὐ μὴ κριθήσεσθε*, φοβούμενος οὐδένα κρίνῃ, ἐν τῇ ἐξετάσει τῶν βεβιωμένων ὡς φύλαξ τῆς ἐντολῆς οὐ κρίνεται· τῆς γὰρ ἑαυτοῦ ἐντολῆς, οὐκ ἐπιλήσμων ὁ ἀψευδέστατος. Οἱ δὲ ἕτεροι δύο τρόποι ἐνταῦθα τὴν συγχώρησιν ἔχουσιν· ὅταν ἐν ἁμαρτίαις τίς ὢν οἰκονομῆται ἐκ τῆς προνοίας

ACPSTDEFHKQ ab ἐξετάσει (l. 22) ACPSTDEFGHQ

1.21–22 Matth. 7.1; Luc. 6.37

1.7 ἠκριβωκότων] ἠκριβηκότων PS 8 δι' ὧν…ἁμαρτημάτων] om. P 9 Ἐπειδὴ] γὰρ add. EKQ 11 μετανοήσῃ] μετανοήσει A 13–14 ὡς εἴρηται] om. DEFHKQ 14 ἐνταῦθα] om. T, post ἁμαρτήσας transp. Q 14–15 ἐνταῦθα…ἀδιαφόρως] om. P 14–15 ἁμαρτήσας…ἀγαθοεργήσας] ἀγαθοεργήσας … ἁμαρτήσας T 17 ἔχομεν] ἔχωμεν A DF$^{2.c.}$, ἐχόμενα EKQ 20 τις] om. F 21 κρίνετε] κρίνεται A T DH (sed ε H$^{s.l.}$) 22 κριθήσεσθε] κριθήσεσθαι A, κριθῆτε Q | κρίνῃ] κρίνει A T DEFHKQ$^{a.c.}$ 24 ἑαυτοῦ] om. PS 26 οἰκονομῆται] GH

Max. 11 ἐνταῦθα] ante μετανοήσῃ transp. μενα Max. 19–20 ὁ^{2}… ἐστὶν] δεύτερος δὲ ριθήσεσθε] ἵνα μὴ κριθῆτε Max. 22 κρίνῃ]



τοῦ πνεύματος ἠκριβωκότων, τέσσαρας εἶναι τρόπους δι' ὧν συγχώρησις γίνεται ἁμαρτημάτων, δύο ἐνταῦθα, καὶ δύο ἐν τῷ μέλλοντι. Ἐπειδὴ οὐκ ἐξικνεῖται ἡ μνήμη ὅλου τοῦ χρόνου μνημονεύειν τὰ σφάλματα ἵνα ὑπὲρ αὐτῶν μετανοήσῃ ὁ ἄνθρωπος ἐνταῦθα, ᾠκονόμησε φιλάνθρωπος ὢν ὁ δεσπότης τῆς φύσεως καὶ ἡμῶν μὴ μετανοούντων τρόπους μετανοίας· ἐν μὲν τῷ μέλλοντι ὡς εἴρηται δύο· ὅταν τις ἀδιαφόρως ἐνταῦθα ἁμαρτήσας καὶ πάλιν ἀδιαφόρως ἀγαθοεργήσας, εἴτε εἰς οἶκτον καὶ συμπάθειαν πρὸς τὸν πλησίον κινηθεὶς ἢ ὅσα ἄλλα φιλανθρωπίας ἔχομεν, ταῦτα ἐν τῷ μέλλοντι αἰῶνι ἐν τῷ καιρῷ τῆς κρίσεως ζυγοστατούμενα ἐφ' ἃ τὴν ῥοπὴν ἕξει, συγχώρησις γίνεται· οὗτος μὲν ὁ εἷς τρόπος· ὁ δὲ δεύτερος ἐστὶν οὗτος· ὅτάν τις ἐν ἁμαρτίαις ἐνεχόμενος, ἀκούων δὲ τοῦ κυρίου λέγοντος *μὴ κρίνετε καὶ οὐ μὴ κριθήσεσθε*, φοβούμενος οὐδένα κρίνῃ, ἐν τῇ ἐξετάσει τῶν βεβιωμένων ὡς φύλαξ τῆς ἐντολῆς οὐ κρίνεται· τῆς γὰρ ἑαυτοῦ ἐντολῆς, οὐκ ἐπιλήσμων ὁ ἀψευδέστατος. Οἱ δὲ ἕτεροι δύο τρόποι ἐνταῦθα τὴν συγχώρησιν ἔχουσιν· ὅταν ἐν ἁμαρτίαις τίς ὢν οἰκονομῆται ἐκ τῆς προνοίας

ACPSTDEFHKQ ab ἐξετάσει (l. 22) ACPSTDEFGHQ

1.21–22 Matth. 7.1; Luc. 6.37

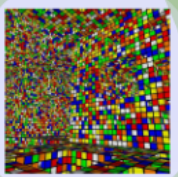
17 ἠκριβωκότων] ἠκριβηκότων PS 8 δι' ὧν…ἁμαρτημάτων] om. P 9 Ἐπειδὴ] γὰρ add. EKQ 11 μετανοήσῃ] μετανοήσει A 13–14 ὡς εἴρηται] om. DEFHKQ 14 ἐνταῦθα] om. T, post ἁμαρτήσας transp. Q 14–15 ἐνταῦθα…ἀδιαφόρως] om. P 14–15 ἁμαρτήσας…ἀγαθοεργήσας] ἀγαθοεργήσας … ἁμαρτήσας T17 ἔχομεν] ἔχωμεν A DF$^{2.c.c}$, ἐχόμενα EKQ 20 τις] om. F21 κρίνετε] κρίνεται A T DH (sed ε H$^{s.l}$) 22 κριθήσεσθε] κριθήσεσθαι A, κριθῆτε Q κρίνῃ] κρίνει A T DEFHKQ$^{s.l}$ 24 ἑαυτοῦ] om. PS26 οἰκονομῆται] οἰκονομεῖται S T DEFGH

Max. 11 ἐνταῦθα] ante μετανοήσῃ transp. Max. 17 ἔχομεν/ἐχόμενα Max. 19–20 ὁ^{2}… ἐστὶν] δεύτερος δὲ Max. 21–22 ἵνα… κριθήσεσθε] ἵνα μὴ κριθῆτε Max. 22 κρίνῃ] κρίνει Max.

Interedition

Formalization

http://dx.doi.org/10.1093/llc/fqu007

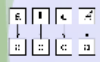


Workflow

1. Tokenization

Gothenburg Model

2. Normalization



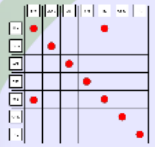
5. Visualization

3. Alignment

4. Analysis

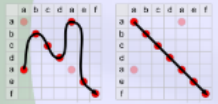
Algorithm

1.



6.

2.



5.

3.

4.

Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project

Ronald Haentjens Dekker
Department of IT R&D, Huygens Institute for the History of the Netherlands, Royal Netherlands Academy of Arts and Sciences, The Netherlands

Dirk van Hulle
Department of Literary Studies, University of Antwerp, Antwerp, Belgium

Gregor Middell
Institut für Deutsche Philologie, Universität Würzburg, Würzburg, Germany

Vincent Neyt
Department of Literary Studies, University of Antwerp, Antwerp, Belgium

Joris van Zundert
Methodology Research Program, Huygens Institute for the History of the Netherlands, Royal Netherlands Academy of Arts and Sciences, The Netherlands

Abstract

Interoperability is the key term within the framework of the European-funded research project Interedition,[1] whose aim is 'to encourage the creators of tools for textual scholarship to make their functionality available to others, and to promote communication between scholars so that we can raise awareness of innovative working methods'. The tools developed by Interedition's 'Prototyping' working group were tested by other research teams, which formulate strategic recommendations. To this purpose, the Centre for Manuscript Genetics (University of Antwerp), the Huygens Institute for the History of the Netherlands (The Hague), and the University of Würzburg have been working together within the framework of Interedition. One of the concrete results of collaboration is the development and fine-tuning of the text collation tool CollateX.[2] In this article, we would like to investigate how the architecture of a

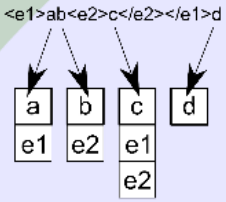
Correspondence:
Joris van Zundert.
Huygens Institute for the History of the Netherlands, Royal Netherlands Academy of Arts and Sciences, PO Box 90754, 2509 LT, The Hague, The Netherlands
Email:
joris.van.zundert@huygens.knaw.nl

Literary and Linguistic Computing © The Author 2014. Published by Oxford University Press on behalf of EADH. All rights reserved. For Permissions, please email: journals.permissions@oup.com
doi:10.1093/llc/fqu007

http://dx.doi.org/10.1093/llc/fqu007

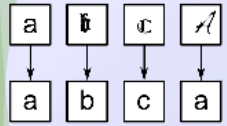
Workflow

1. Tokenization



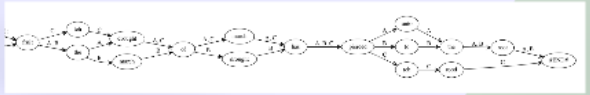
Gothenburg Model

2. Normalization

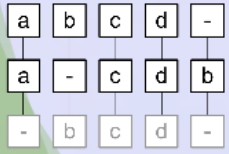


5. Visualization

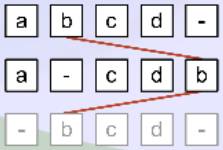
A	The	quick	brown	fox jumped over the	lazy	dog.
B	The	-	brown	fox jumped over the	-	dog.
C	The	bad	-	fox jumped over the	lazy	dog.



3. Alignment

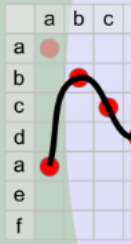


4. Analysis

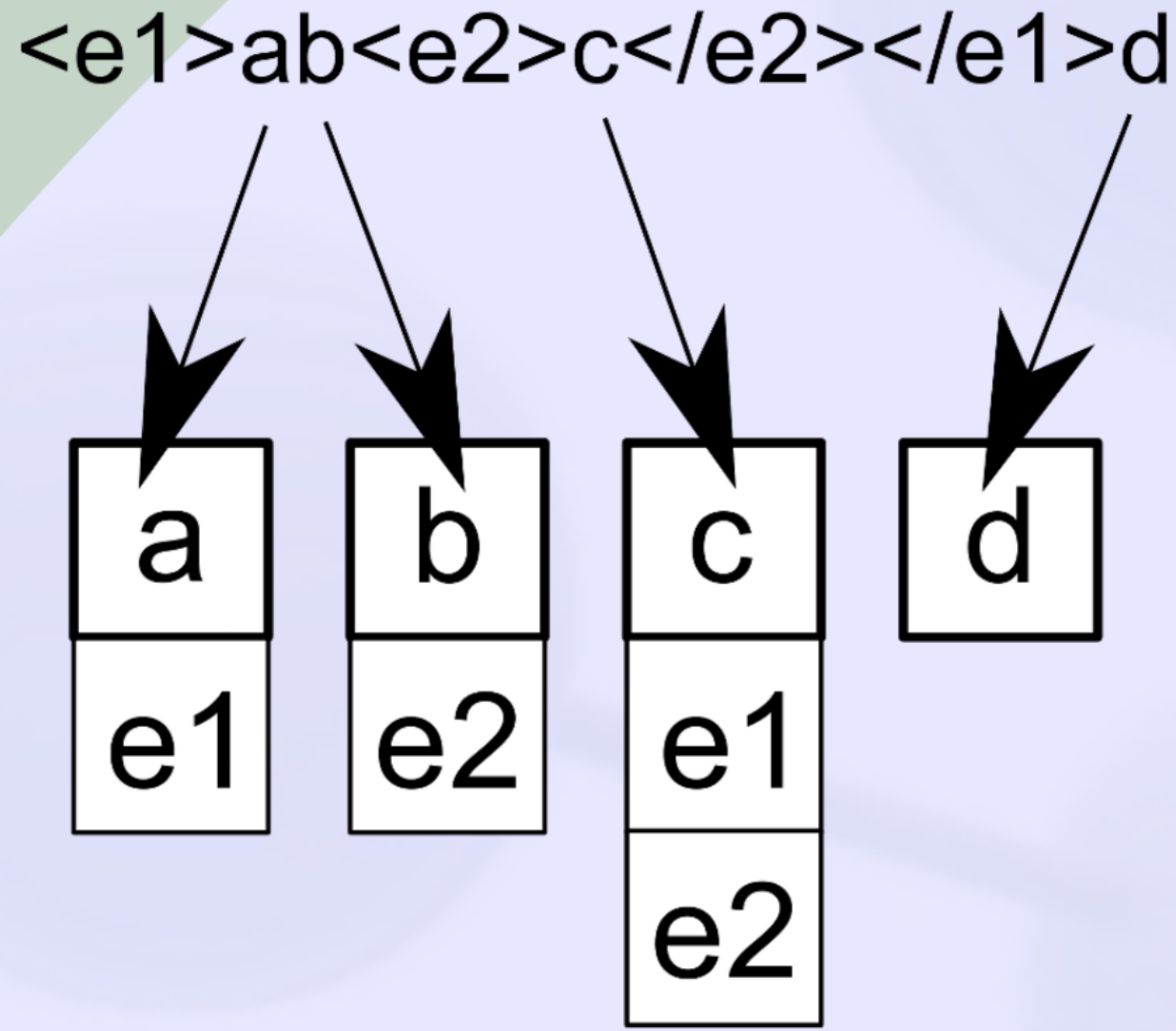


1.

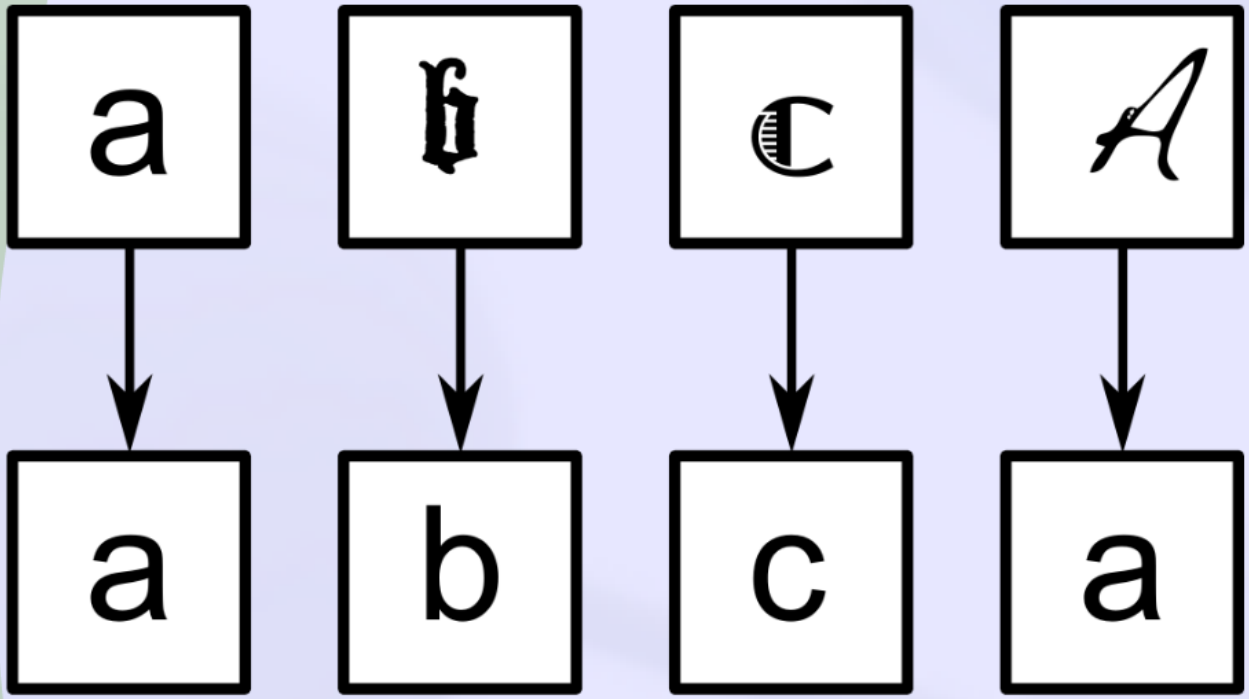
2.



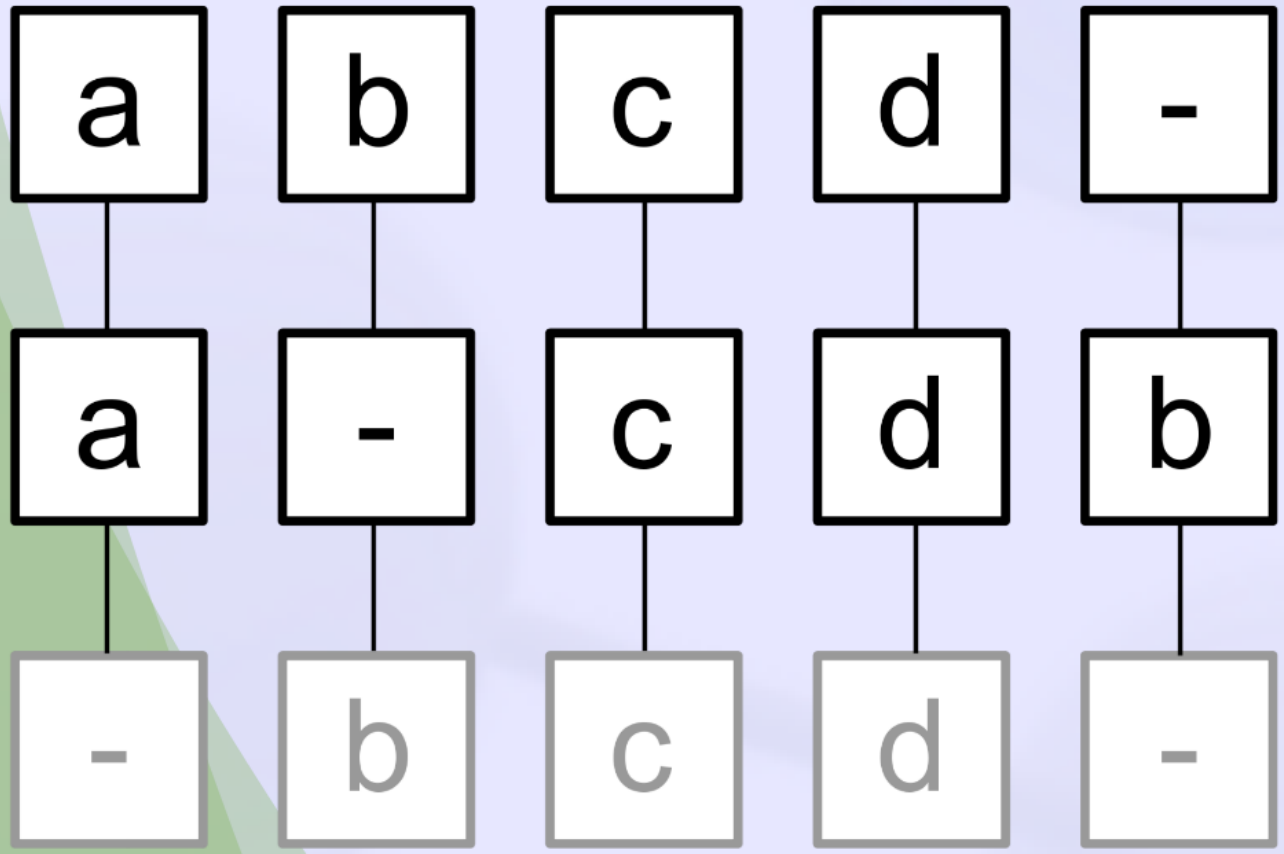
1. Tokenization



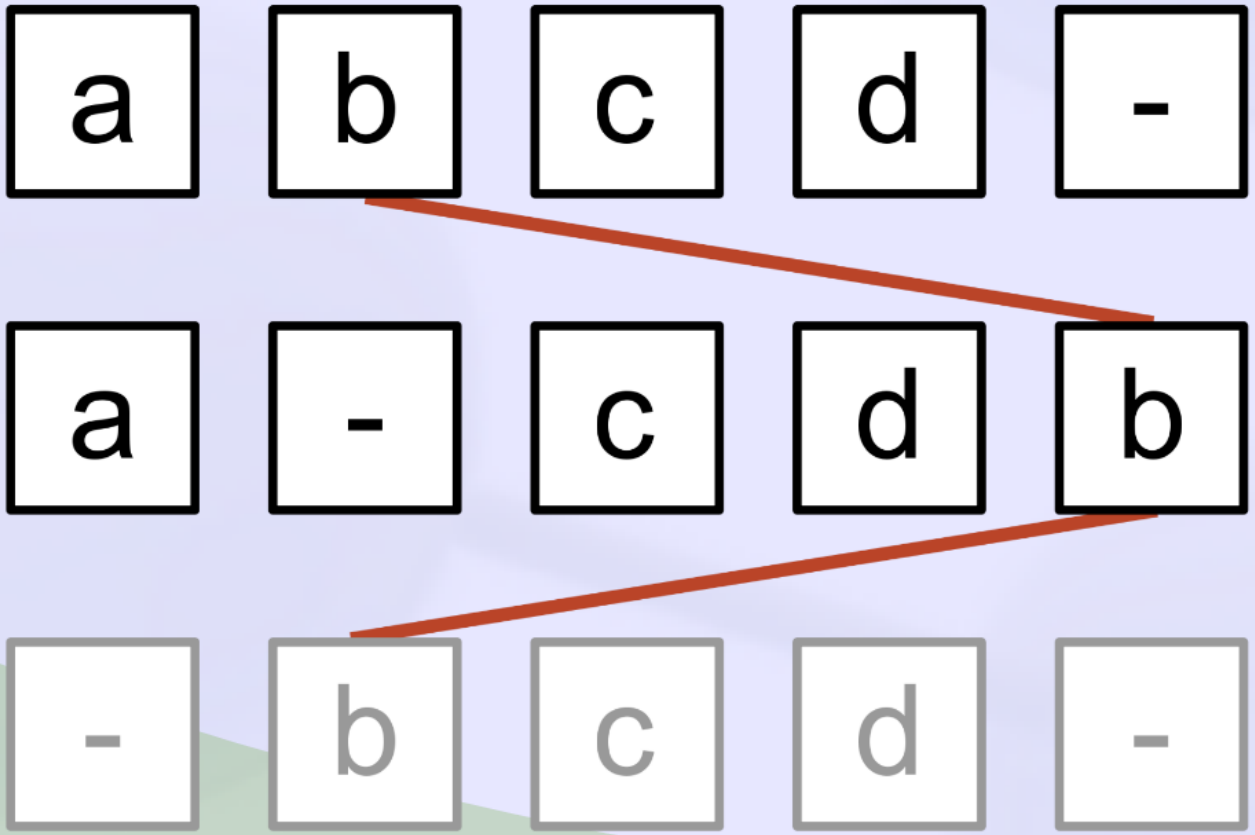
2. Normalization



3. Alignment

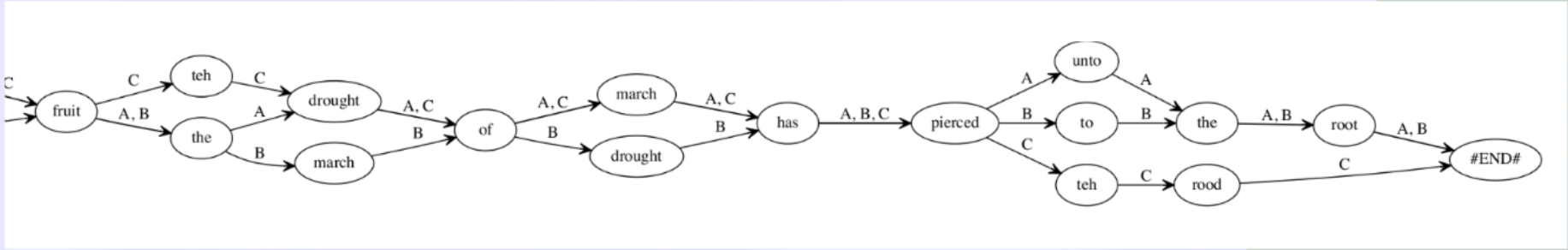


4. Analysis



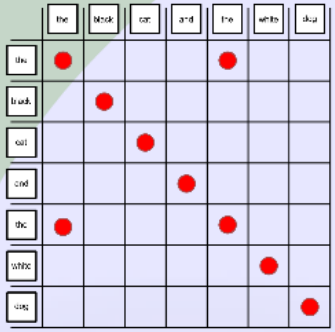
5. Visualization

A	The	quick	brown	fox jumped over the	lazy	dog.
B	The	-	brown	fox jumped over the	-	dog.
C	The	bad	-	fox jumped over the	lazy	dog.

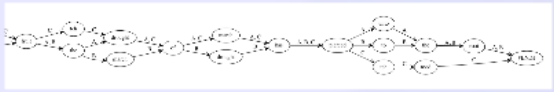


Algorithm

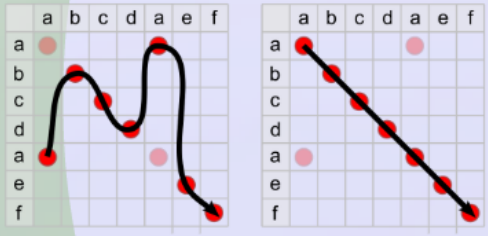
1.



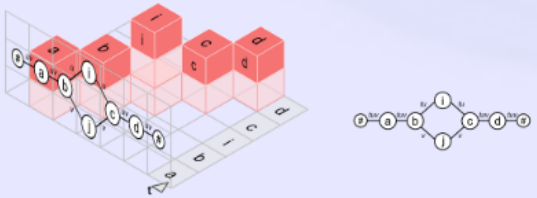
6.



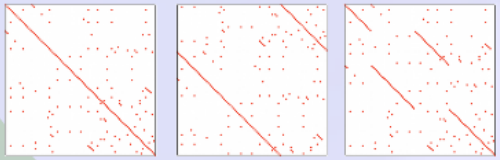
2.



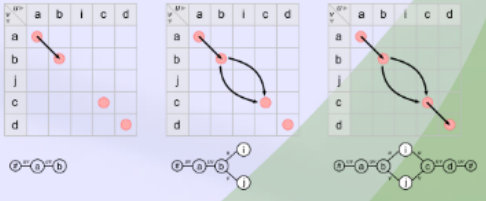
5.

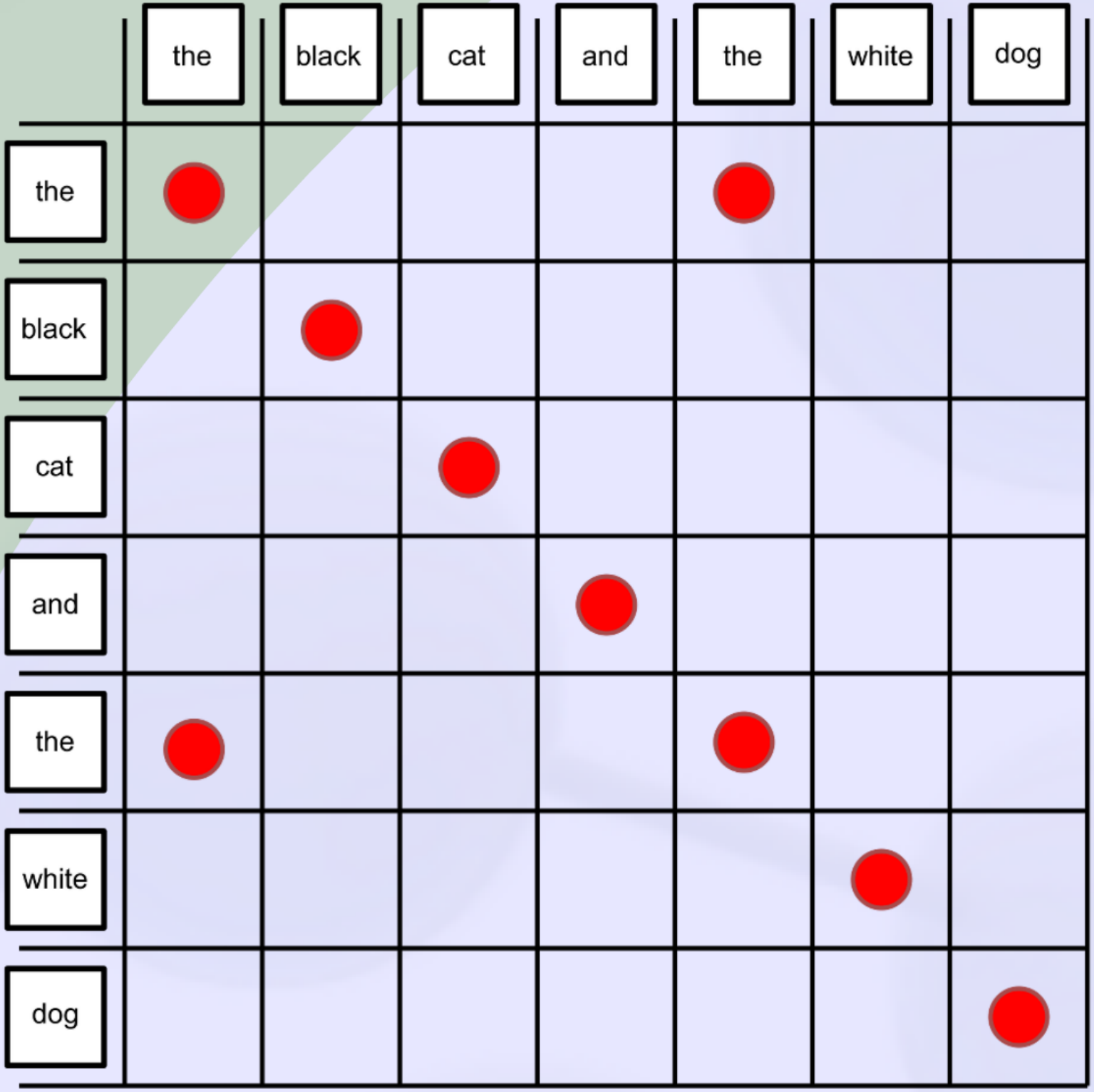


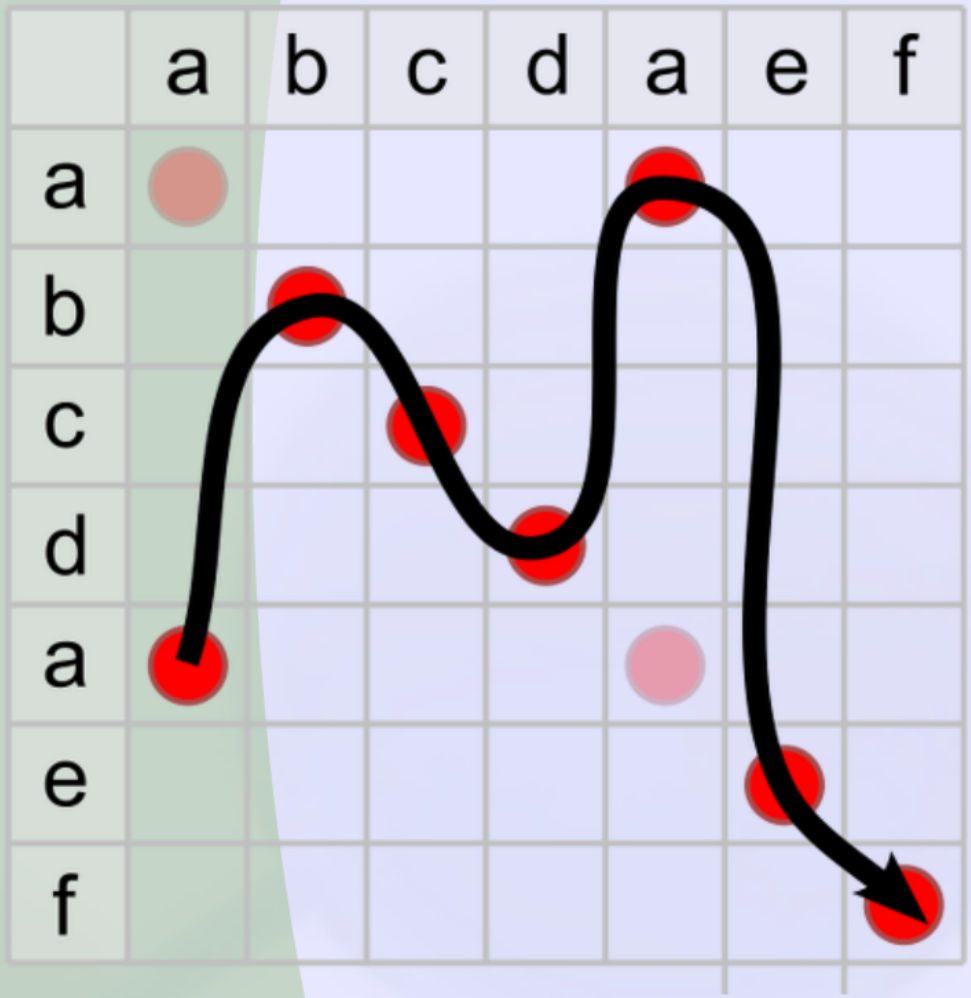
3.

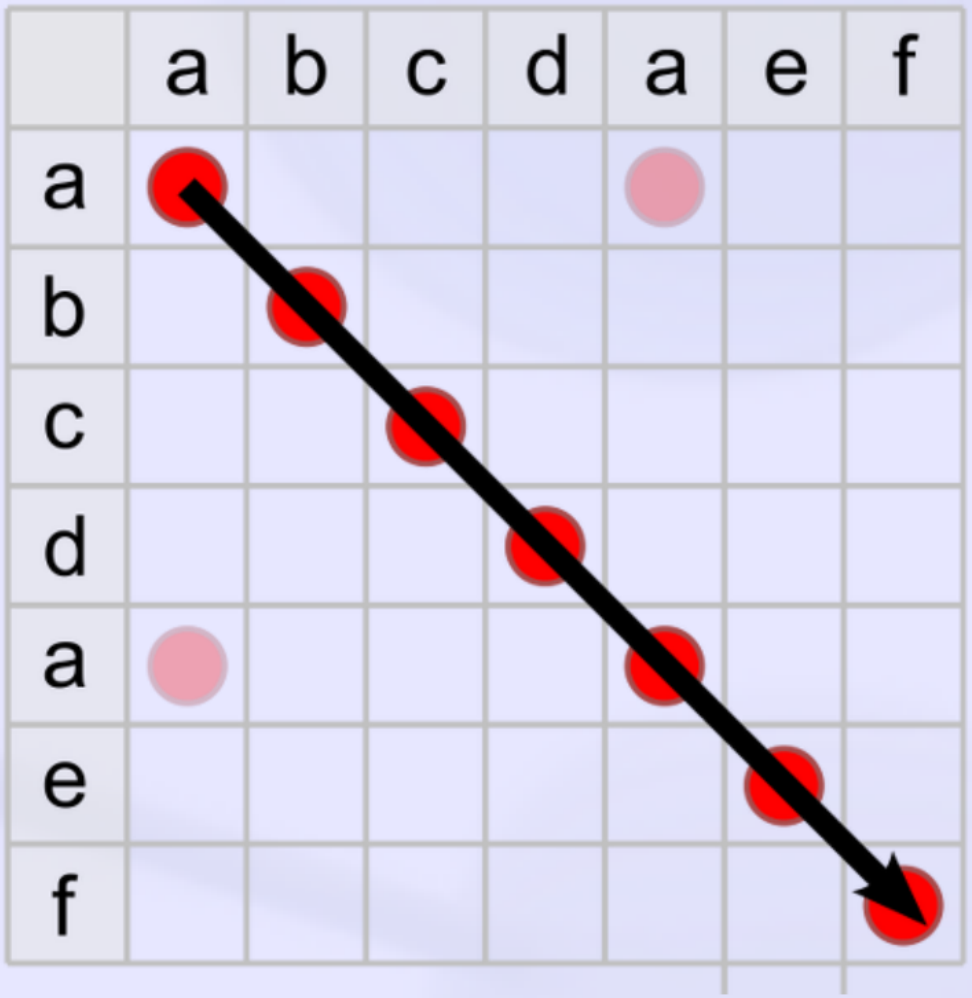


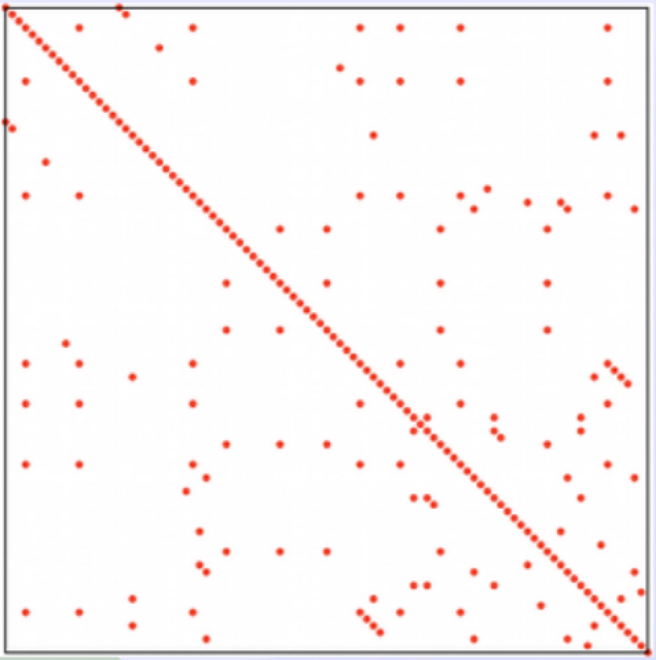
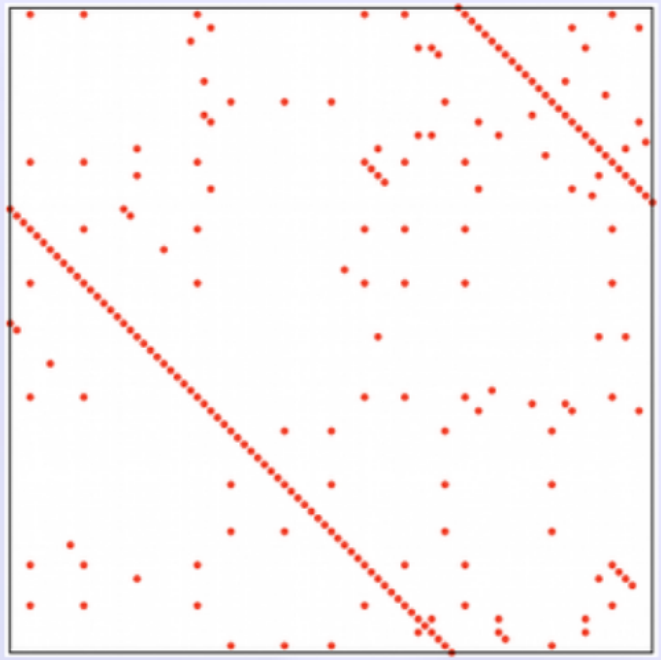
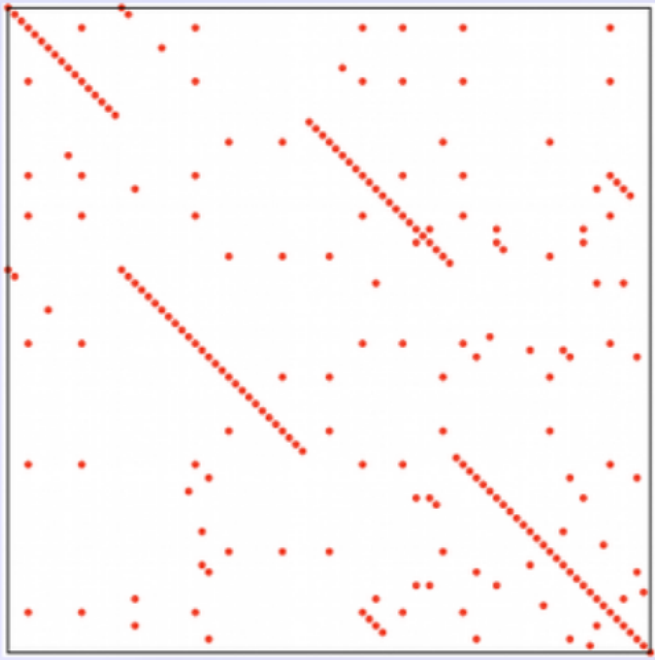
4.

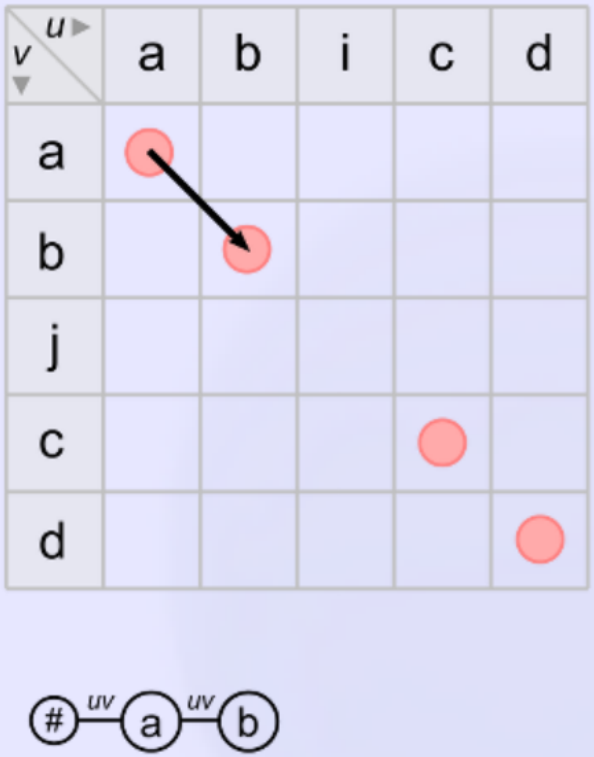



the
black
cat
and
the
white
dog
the
black
cat
and
the
white
dog

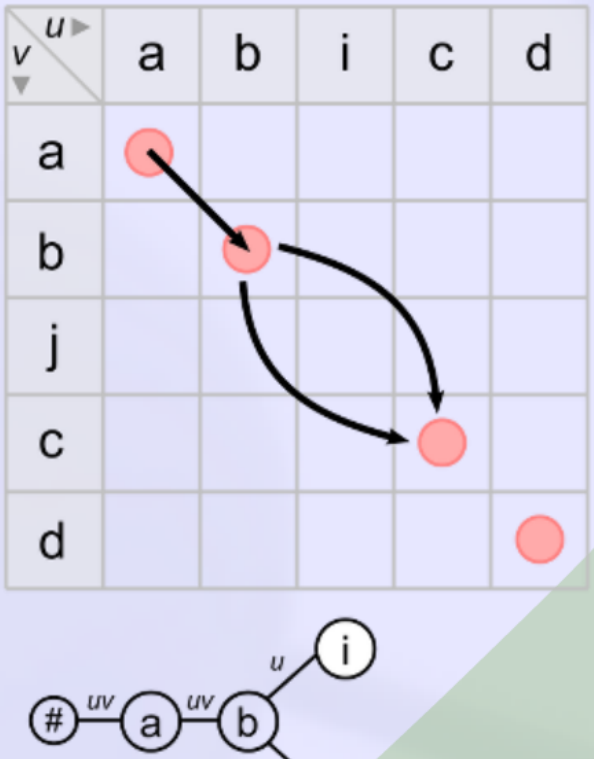

a b c d a e f
a
b
c
d
a
e
f

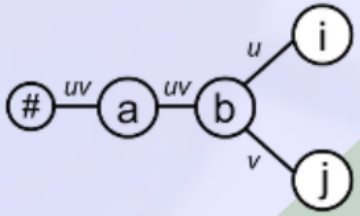

a b c d a e f
a
b
c
d
a
e
f

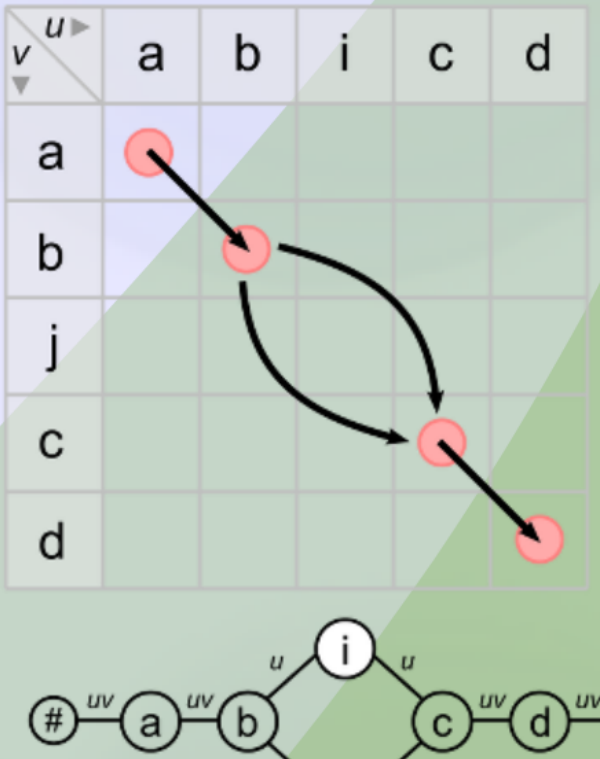




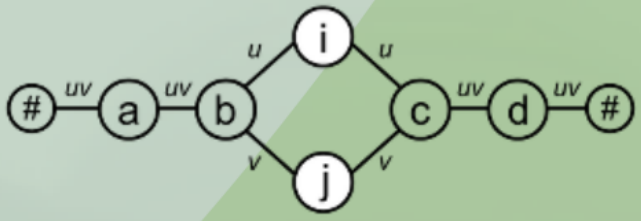

u
v
a
b
i
c
d
a
b
j
c
d

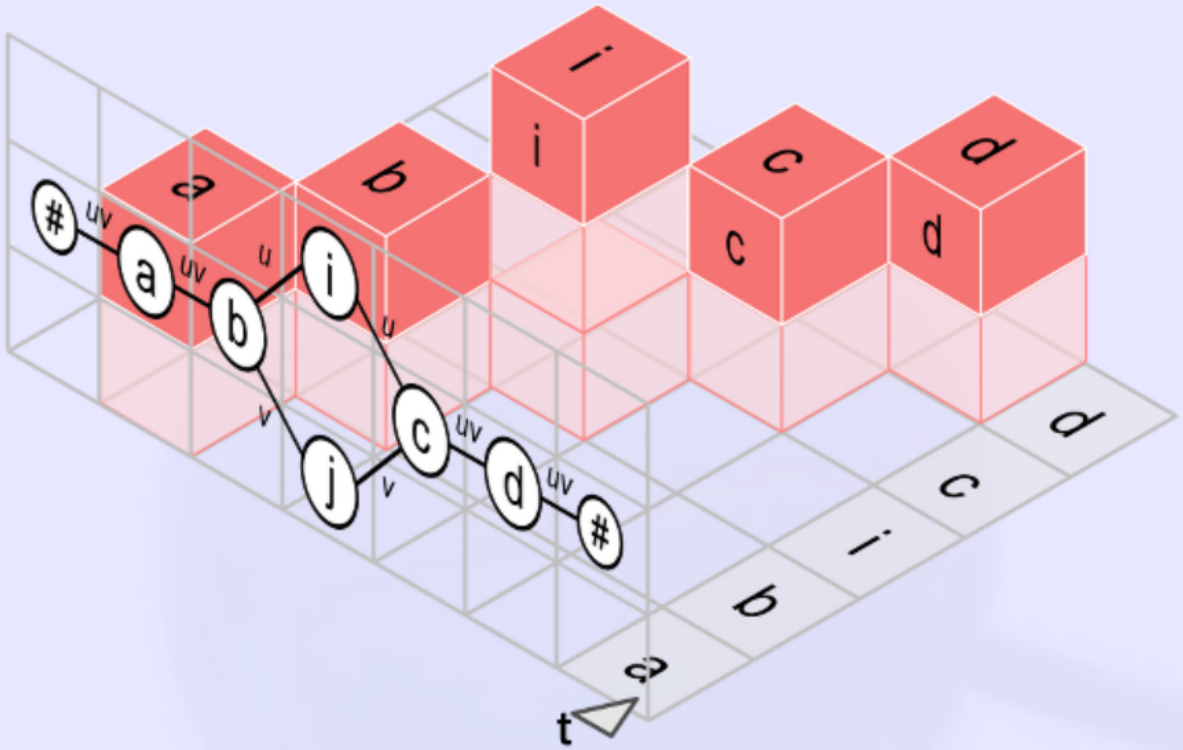
#
uv
a
uv
b

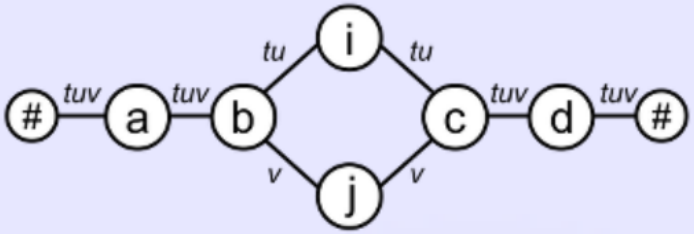

u
v
a
b
i
c
d
a
b
j
c
d

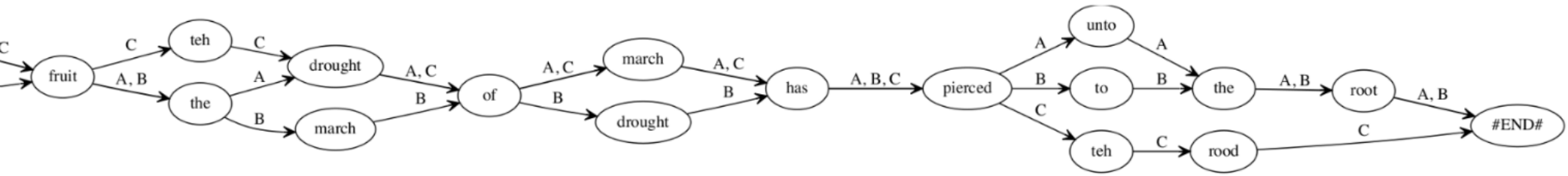

#
uv
a
uv
b
u
i
v
j

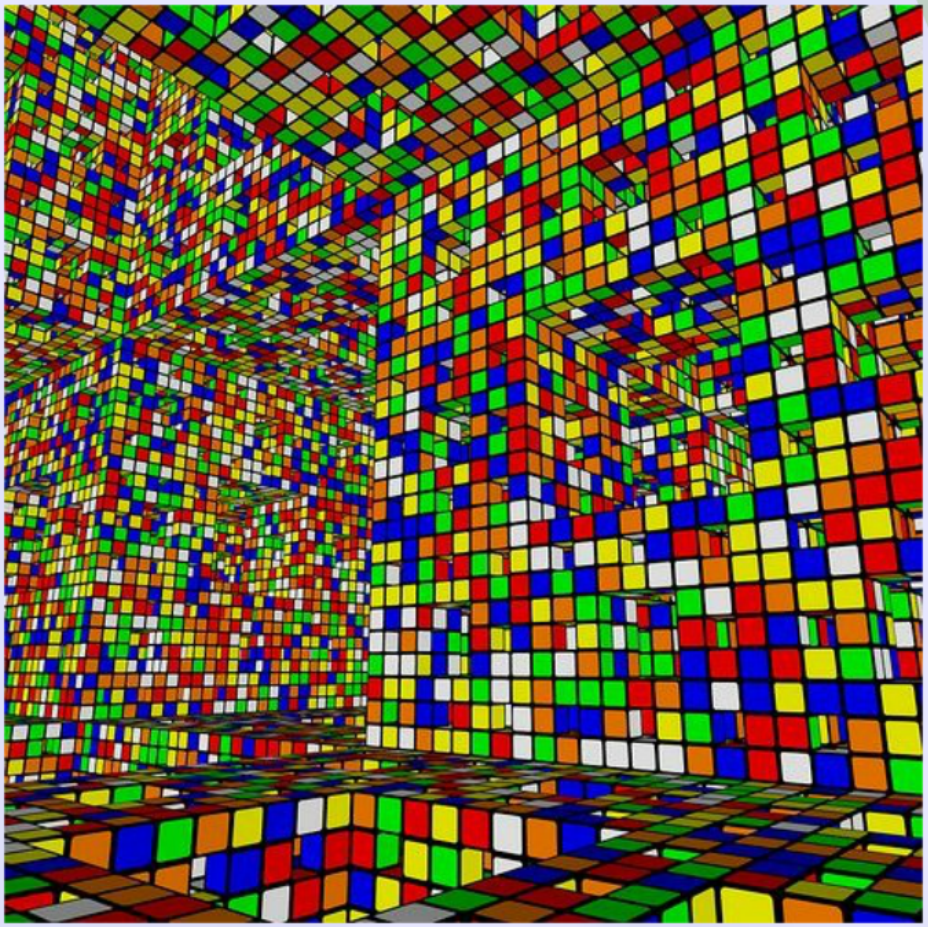

u
v
a
b
i
c
d
a
b
j
c
d


#
uv
a
uv
b
u
i
u
v
j
v
c
uv
d
uv
#


a
b
i
c
d
#
uv
uv
u
u
v
uv
v
uv
#
t


#
tuv
a
tuv
b
tu
i
tu
v
j
v
c
tuv
d
tuv
#


C
fruit
C
A, B
teh
the
C
A
B
drought
march
A, C
B
of
A, C
B
march
drought
A, C
B
has
A, B, C
pierced
A
B
C
unto
to
teh
A
B
C
the
rood
A, B
C
root
A, B
C
#END#



The Problem II

Judits ieeste gaet hier an
Van liue wijf van herten man
UErstaet Nu wel Na cyrus doot.
Wart een arfaxat here groot.
Die ueste gathamus de stede
Die staet in dat lant van mede
Ommedat hi wilde moghendelike

Aa

Judiths geeste gaet hier an.
Van liue wijf van herten man.
Verstaet mi wel na cirus dod.
Ward een arfaxat grod.
Ende veste ecbatanis die stede.
Die staet int landscap ende heet mede.
Omdat hi wilde moghenlike.

Judits ieeste gaet hier an
Van liue wijf van herten man
UErstaet Nu wel Na cyrus doot.
Wart een arfaxat here groot.
Die ueste gathamus de stede
Die staet in dat lant van mede
Ommedat hi wilde moghendelike

Judiths geeste gaet hier an.
Van liue wijf van herten man.
Verstaet mi wel na cirus dod.
Ward een arfaxat grod.
Ende veste ecbatanis die stede.
Die staet int landscap ende heet mede.
Omdat hi wilde moghenlike.

The Problem II

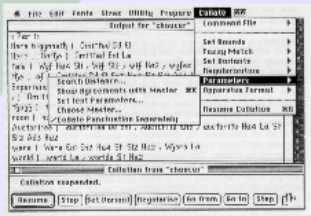
Judits ieeste gaet hier an
Van liue wijf van herten man
UErstaet Nu wel Na cyrus doot.
Wart een arfaxat here groot.
Die ueste gathamus de stede
Die staet in dat lant van mede
Ommedat hi wilde moghendelike

Aa

Judiths geeste gaet hier an.
Van liue wijf van herten man.
Verstaet mi wel na cirus dod.
Ward een arfaxat grod.
Ende veste ecbatanis die stede.
Die staet int landscap ende heet mede.
Omdat hi wilde moghenlike.

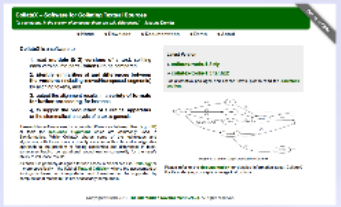
Tools

Collate



CollateX

http://collatex.net/



Juxta

http://www.juxtacommons.org/

diff/wdiff

http://www.gnu.org/software/wdiff/

TEIComparator

http://tei-comparator.sourceforge.net/

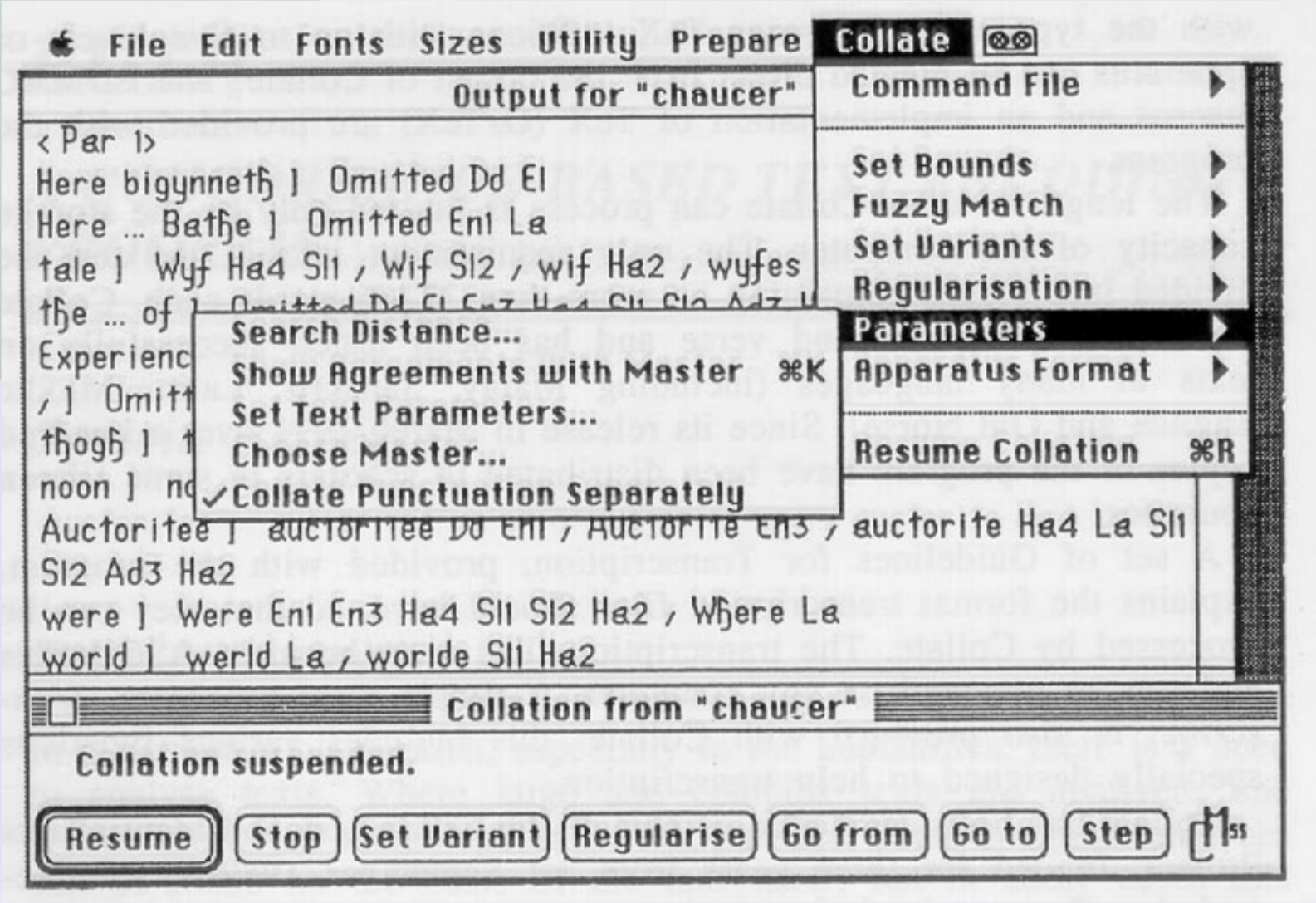
TuStep

http://tustep.wikispaces.com/TUSTEP-Wiki

NMerge

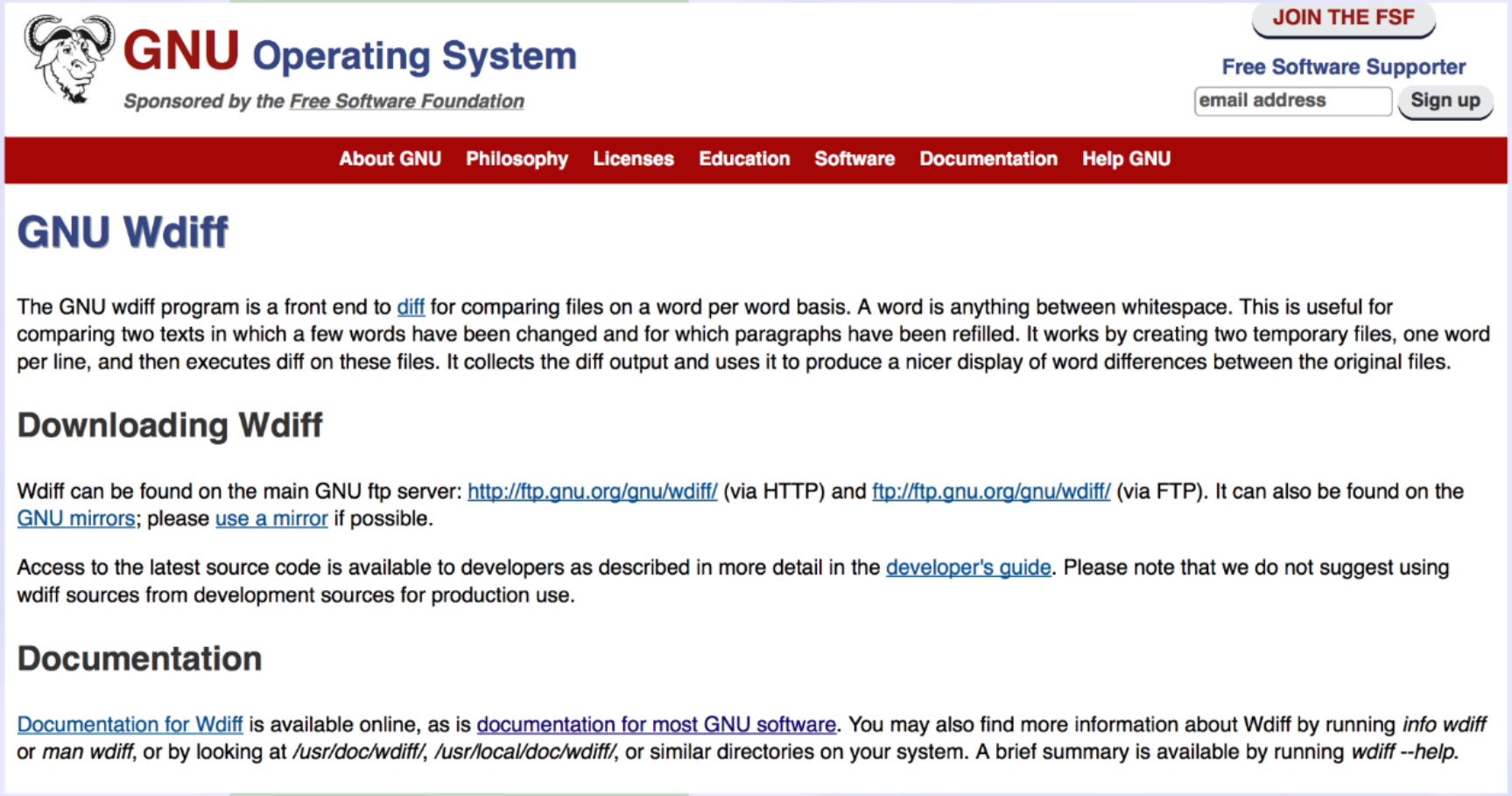
https://code.google.com/p/multiversiondocs/wiki/NMerge

Collate


File Edit Fonts Sizes Utility Prepare Collate
Output for "chaucer"
Command File
Set Bounds
Fuzzy Match
Set Variants
Regularisation
Parameters
Apparatus Format
Resume Collation ⌘R
Search Distance...
Show Agreements with Master ⌘K
Set Text Parameters...
Choose Master...
✓Collate Punctuation Separately
< Par 1>
Here bigynneth] Omitted Dd El
Here ... Bathe] Omitted En1 La
tale] Wyf Ha4 Sl1 , Wif Sl2 , wif Ha2 , wyfes
Auctoritee] auctoritee Dd En1 , Auctorite En3 , auctorite Ha4 La Sl1
Sl2 Ad3 Ha2
were] Were En1 En3 Ha4 Sl1 Sl2 Ha2 , Where La
world] werld La , worlde Sl1 Ha2
Collation from "chaucer"
Collation suspended.
Resume
Stop
Set Variant
Regularise
Go from
Go to
Step

diff/wdiff

http://www.gnu.org/software/wdiff/



GNU Operating System

Sponsored by the Free Software Foundation

JOIN THE FSF

Free Software Supporter

email address Sign up

About GNU Philosophy Licenses Education Software Documentation Help GNU

GNU Wdiff

The GNU wdiff program is a front end to diff for comparing files on a word per word basis. A word is anything between whitespace. This is useful for comparing two texts in which a few words have been changed and for which paragraphs have been refilled. It works by creating two temporary files, one word per line, and then executes diff on these files. It collects the diff output and uses it to produce a nicer display of word differences between the original files.

Downloading Wdiff

Wdiff can be found on the main GNU ftp server: http://ftp.gnu.org/gnu/wdiff/ (via HTTP) and ftp://ftp.gnu.org/gnu/wdiff/ (via FTP). It can also be found on the GNU mirrors; please use a mirror if possible.

Access to the latest source code is available to developers as described in more detail in the developer's guide. Please note that we do not suggest using wdiff sources from development sources for production use.

Documentation

Documentation for Wdiff is available online, as is documentation for most GNU software. You may also find more information about Wdiff by running *info wdiff* or *man wdiff*, or by looking at */usr/doc/wdiff/*, */usr/local/doc/wdiff/*, or similar directories on your system. A brief summary is available by running *wdiff --help*.

TuStep

http://tustep.wikispaces.com/TUSTEP-Wiki

TUSTEP

guest | Join | Help | Sign In

Search

Home

Was ist TUSTEP?
Downloads
Einführungen
Links
Tutorial: Erster Aufruf
Tutorial: Erste Schritte
Tutorial: Grundlagen
Mögliche Probleme
TUSTEP Lösungen und Tipps

Was ist TUSCRIPT?
Tutorial: Erste Schritte
TUSCRIPT Lösungen und Tipps

Was ist TXSTEP?
TXSTEP Erste Schritte
TXSTEP Grundlagen
TXSTEP Lösungen und Tipps

Präsentationen
Workshops
Videos
Projekte

Anleitung für Autoren
Muster für Seite oder Post
Impressum

Willkommen zum TUSTEP-Wiki!

Dieses Wiki bietet **Informationen und Materialien** zum "Tübinger System von Textverarbeitungs-Programmen" **TUSTEP**.

Ziel ist eine **Sammlung von Problemen, Lösungen und praktischen Beispielen** für alle, die mit TUSTEP arbeiten, für Anfänger ebenso wie für fortgeschrittene Anwender.

Quelle: ASCII Art

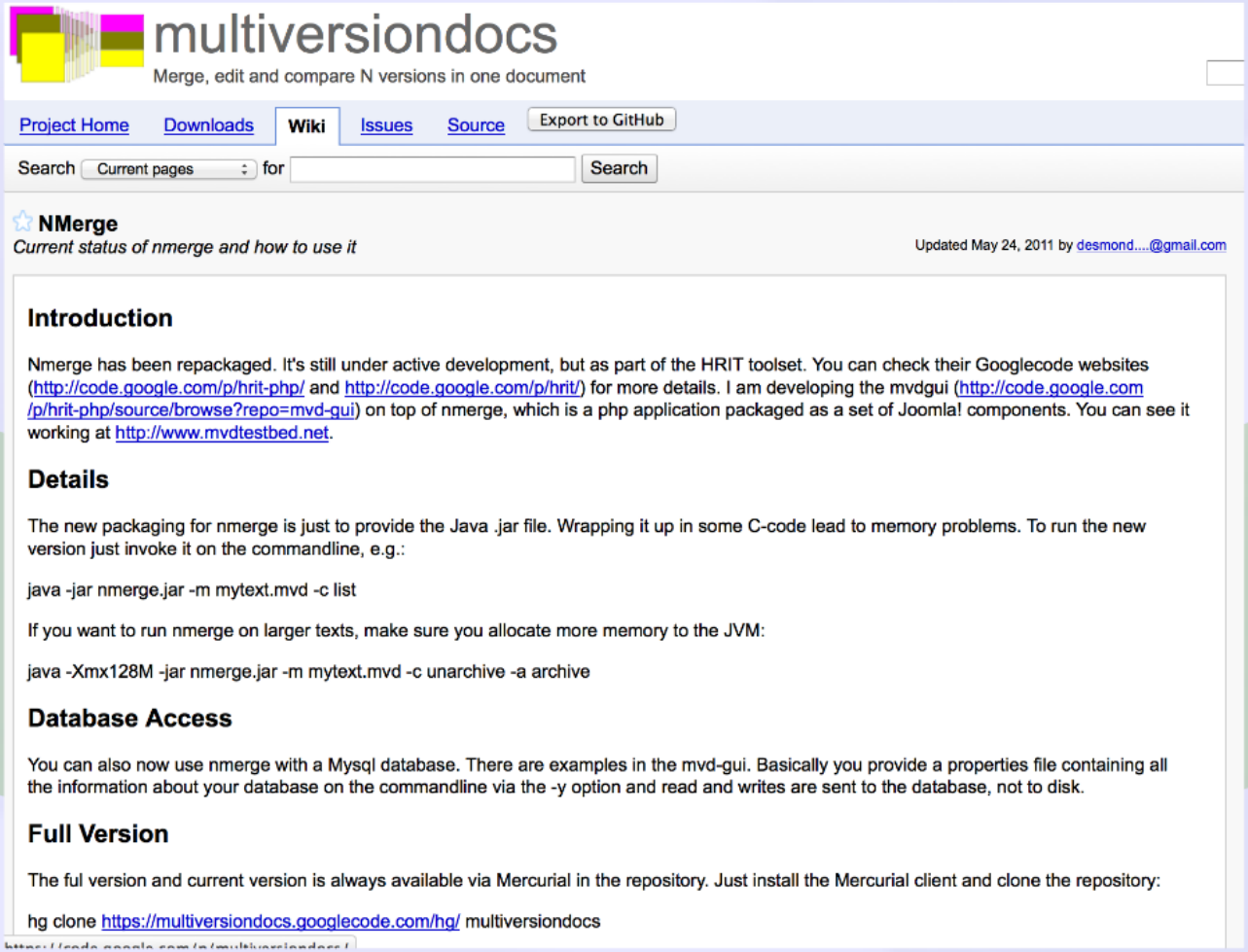
Welcome to the TUSTEP Wiki!

This Wiki offers **information** and **resources** related to the "Tubingen System of Text Processing Tools" **TUSTEP**.

The aim is a **collection of problems, solutions, and practical examples** for anyone working with TUSTEP, for beginners as well as for experienced users.

NMerge

https://code.google.com/p/multiversiondocs/wiki/NMerge



multiversiondocs

Merge, edit and compare N versions in one document

Project Home | Downloads | **Wiki** | Issues | Source | Export to GitHub

Search Current pages for Search

NMerge

Current status of nmerge and how to use it

Updated May 24, 2011 by desmond....@gmail.com

Introduction

Nmerge has been repackaged. It's still under active development, but as part of the HRIT toolset. You can check their Googlecode websites (http://code.google.com/p/hrit-php/ and http://code.google.com/p/hrit/) for more details. I am developing the mvdgui (http://code.google.com/p/hrit-php/source/browse?repo=mvd-gui) on top of nmerge, which is a php application packaged as a set of Joomla! components. You can see it working at http://www.mvdtestbed.net.

Details

The new packaging for nmerge is just to provide the Java .jar file. Wrapping it up in some C-code lead to memory problems. To run the new version just invoke it on the commandline, e.g.:

java -jar nmerge.jar -m mytext.mvd -c list

If you want to run nmerge on larger texts, make sure you allocate more memory to the JVM:

java -Xmx128M -jar nmerge.jar -m mytext.mvd -c unarchive -a archive

Database Access

You can also now use nmerge with a Mysql database. There are examples in the mvd-gui. Basically you provide a properties file containing all the information about your database on the commandline via the -y option and read and writes are sent to the database, not to disk.

Full Version

The ful version and current version is always available via Mercurial in the repository. Just install the Mercurial client and clone the repository:

hg clone https://multiversiondocs.googlecode.com/hg/ multiversiondocs

TEIComparator

http://tei-comparator.sourceforge.net/

TEI-Comparator

About TEI-Comparator

The TEI-Comparator is a text-comparison engine designed to compare two XML files where items of a paragraph-like granularity have been moved around, split up or otherwise reorganized. It is called the 'TEI-Comparator' because originally it was designed to compare two large XML files that follow the Guidelines of the Text Encoding Initiative (TEI). However, the TEI-Comparator does not specifically require TEI XML and should work with any XML as long as the units being compared are of a similar paragraph-like size. It is a database-backed (MySQL or HSQL or indeed any database that is supported by Hibernate) java web application built using the Google Web Toolkit API.

Links

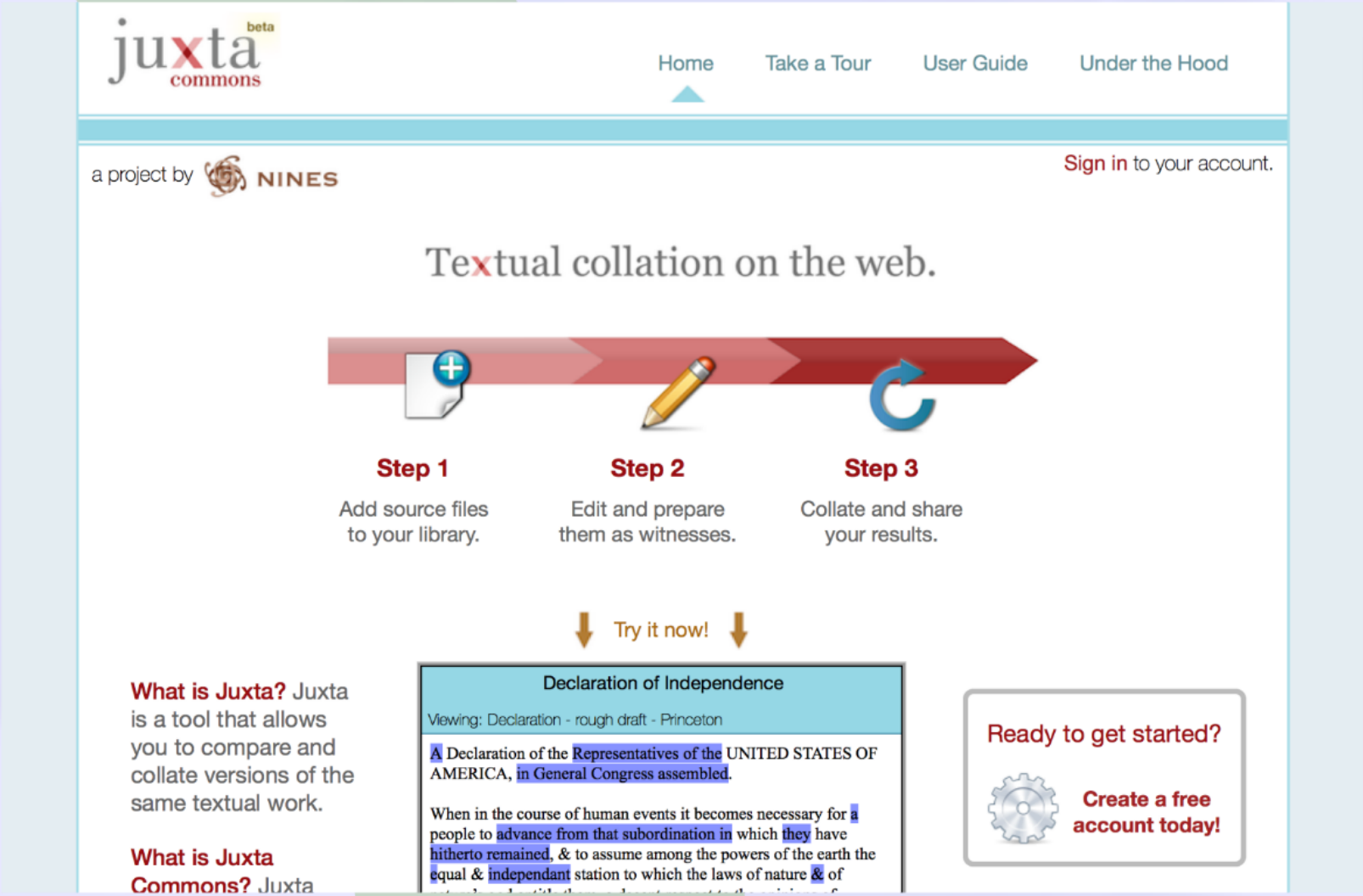
- Project Summary Page
- Downloads Page
- Project Support Page
- Project Documentation Wiki
- Help Forum
- Open Discussion Forum
- Bug Tracker
- Feature Requests
- Screenshots
- Develop (including SVN repository information)

Contact

The TEI-Comparator is maintained by Arno Mittelbach and James Cummings. For questions, comments, or suggestions regarding the TEI-Comparator you can contact Arno at firstname.lastname[at]kom.tu-darmstadt.de and James at firstname.lastname[at]oucs.ox.ac.uk.

Juxta

http://www.juxtacommons.org/



CollateX

http://collatex.net/

CollateX – Software for Collating Textual Sources

"In a language, in the system of language, there are only differences." – Jacques Derrida

Fork me on GitHub

» Home » Download » Documentation » Demo » About

CollateX is a software to

1. read **multiple (≥ 2) versions of a text**, splitting each version into parts (**tokens**) to be compared,
2. **identify similarities of and differences between the versions** (including moved/transposed segments) by aligning tokens, and
3. output the alignment results in a **variety of formats for further processing**, for instance
4. to support **the production of a critical apparatus** or the stemmatical analysis of a text's genesis.

It resembles software used to compute differences between files (e.g. **diff**) or tools for **sequence alignment** which are commonly used in Bioinformatics. While CollateX shares some of the techniques and algorithms with those tools, it mainly aims for a flexible and configurable approach to the problem of finding similarities and differences in texts, sometimes trading computational soundness or complexity for the user's ability to influence results.

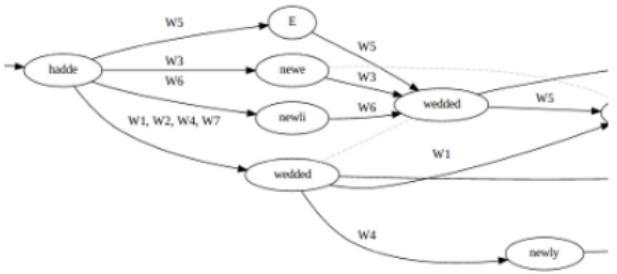
As such it is primarily designed for use cases in disciplines like **Philology** or – more specifically – the field of **Textual Criticism** where the assessment of findings is based on interpretation and therefore can be supported by computational means but is not necessarily computable.

Latest Version

» **collatex-tools-1.3.zip**

» **collatex-tools-1.3.tar.bz2**

For alternative packages and license terms, please read the **download section**.



Snippet of a Variant Graph produced by CollateX

Please refer to the **documentation** for detailed information about CollateX like its underlying concepts or usage instructions.

Copyright © 2010-2013 **The Interedition Development Group**. All rights reserved.

Automated Collation

Interedition

Tara L. Andrews
tara.andrews@kps.unibe.ch

Leif-Jöran Olsson
leif-joran.olsson@svenska.gu.se

David J. Birnbaum
djbpitt@gmail.com

Joris J. van Zundert
joris.van.zundert@huygens.knaw.nl

Ronald Haentjens Dekker
ronald.dekker@huygens.knaw.nl