Wuppertal 2020-02-06

Ich meinte, sie sagte, er schrieb, dass er dachte...

Transcriptions as recordings of contextualized document changes

Gioele Barabucci
Institute of design

NTNU | Norwegian University of Science and Technology

Text genetics, collaborative editing, document changes and versioning

Gioele Barabucci

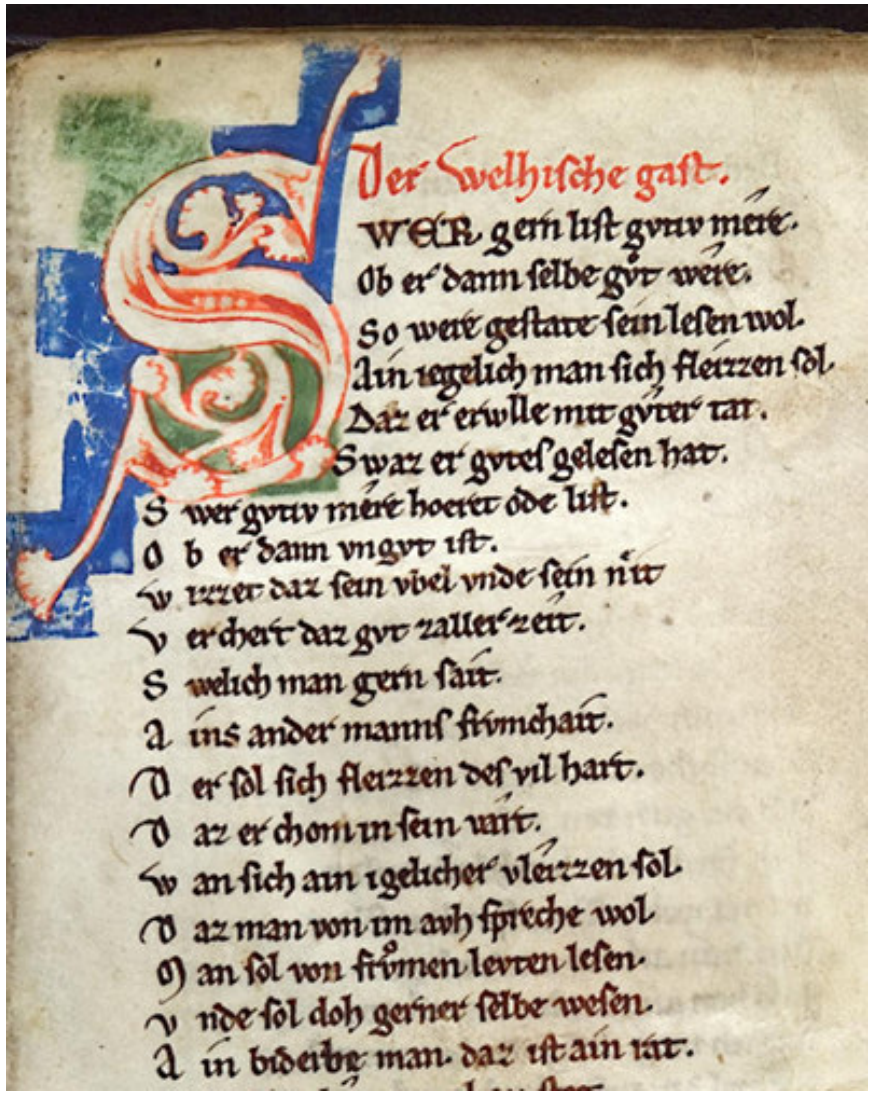
Universität zu Köln

(1)

Chapter 1

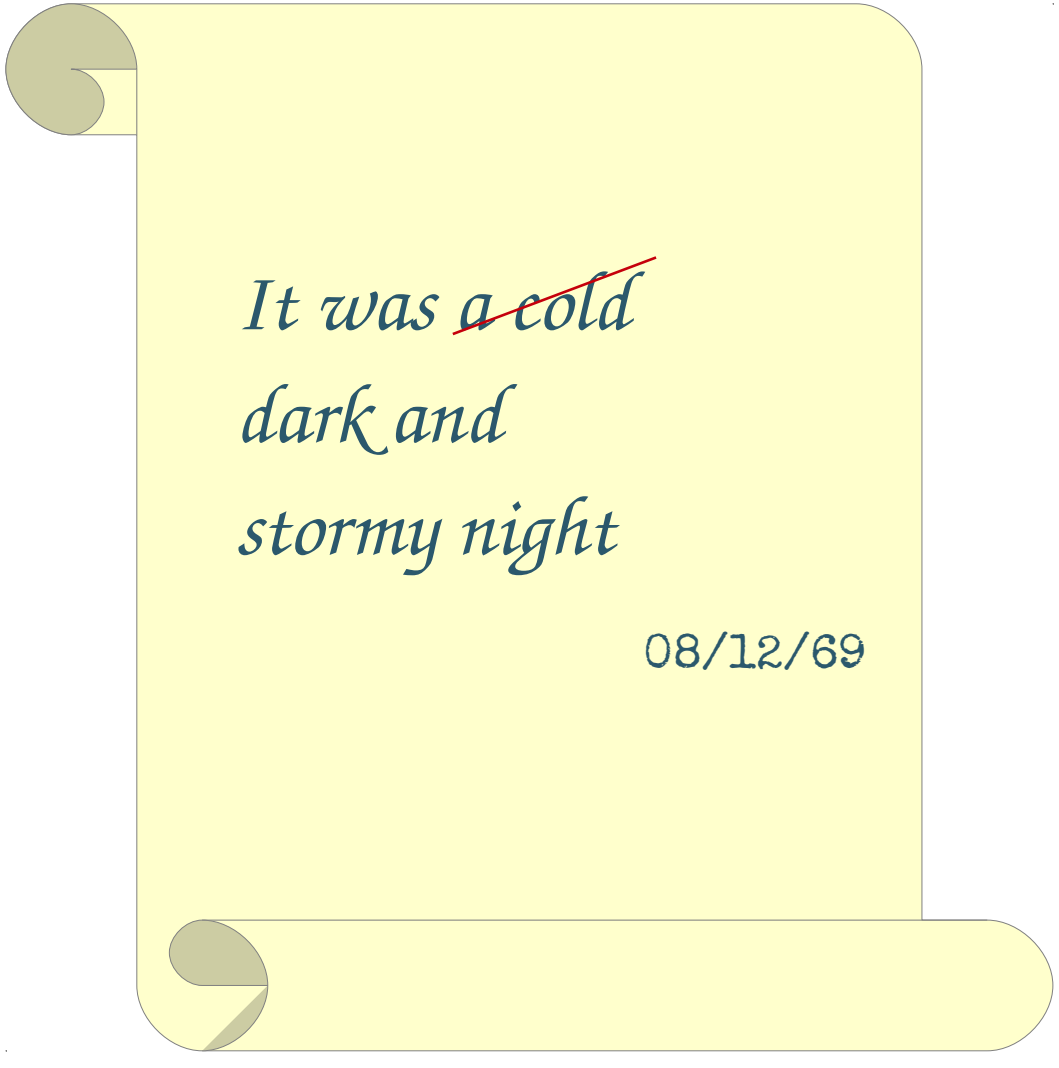
Vaughan died yesterday in his last car crash. During ~~the~~ our ~~time I had known him~~ he had ~~been~~ in many crashes, but this was his only true accident. Driven on a collision course towards the limousine of the film actress whom he had dreamed of killing for so many months, his car jumped the rails of the flyover and plunged through the roof of a bus filled with airline passengers. The crushed bodies of dead ~~[illegible]~~ like a heamorrhage of the sun, still lay across the ~~twisted~~ seats when I pushed my way through the police engineers an hour later. Holding the arm of her chauffeur, ~~[illegible]~~ Taylor, stood ~~by~~ the revolving light of an ambulance, a gloved hand to her throat. ~~Was she aware of the death Vaughan had planned for her?~~

Vaughan ~~had~~ thought of nothing else but her death during the last weeks of his life, a coronation of wounds he had devised with the care of an earl marshall. The walls of his ~~[illegible]~~ apartment were covered with ~~[illegible]~~ photographs he had taken as she left her hotel [illegible] his room long from

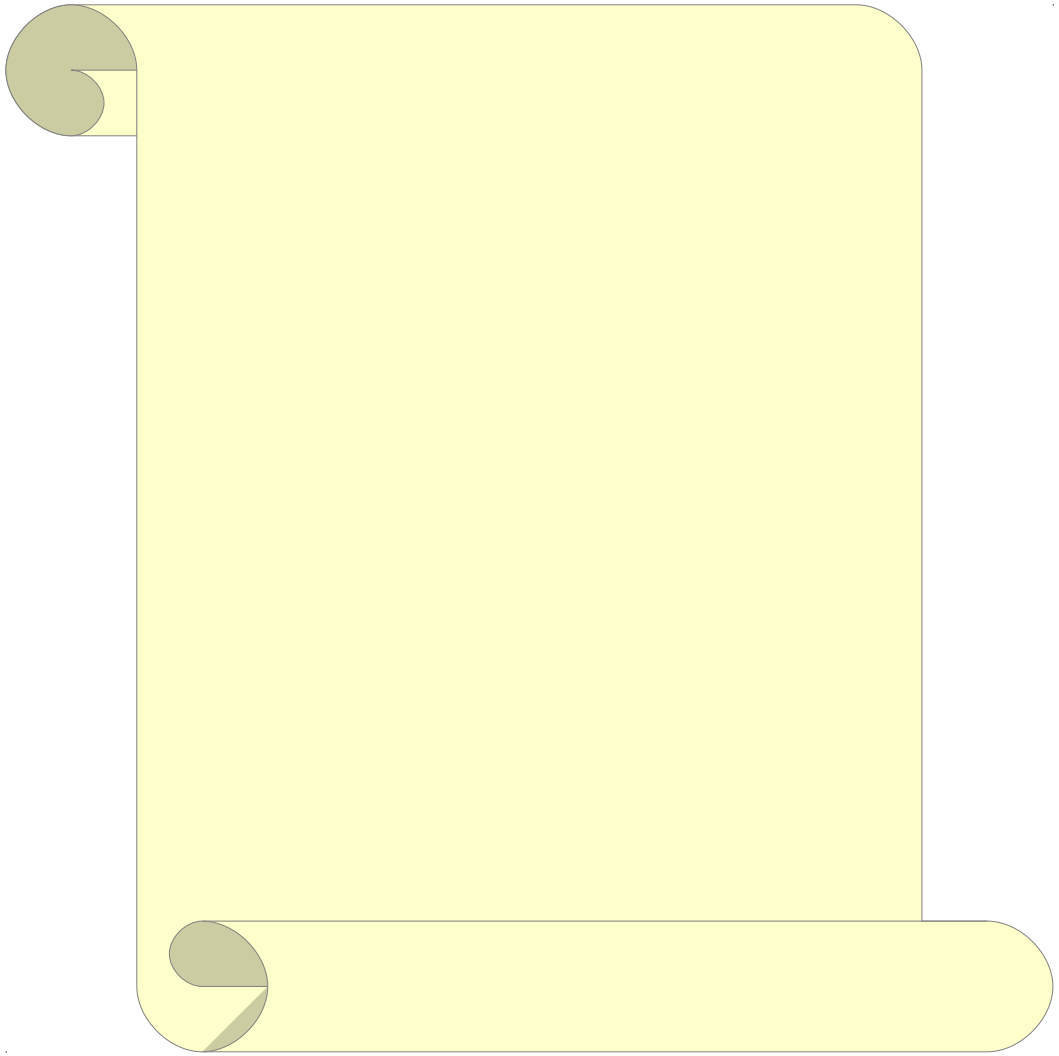


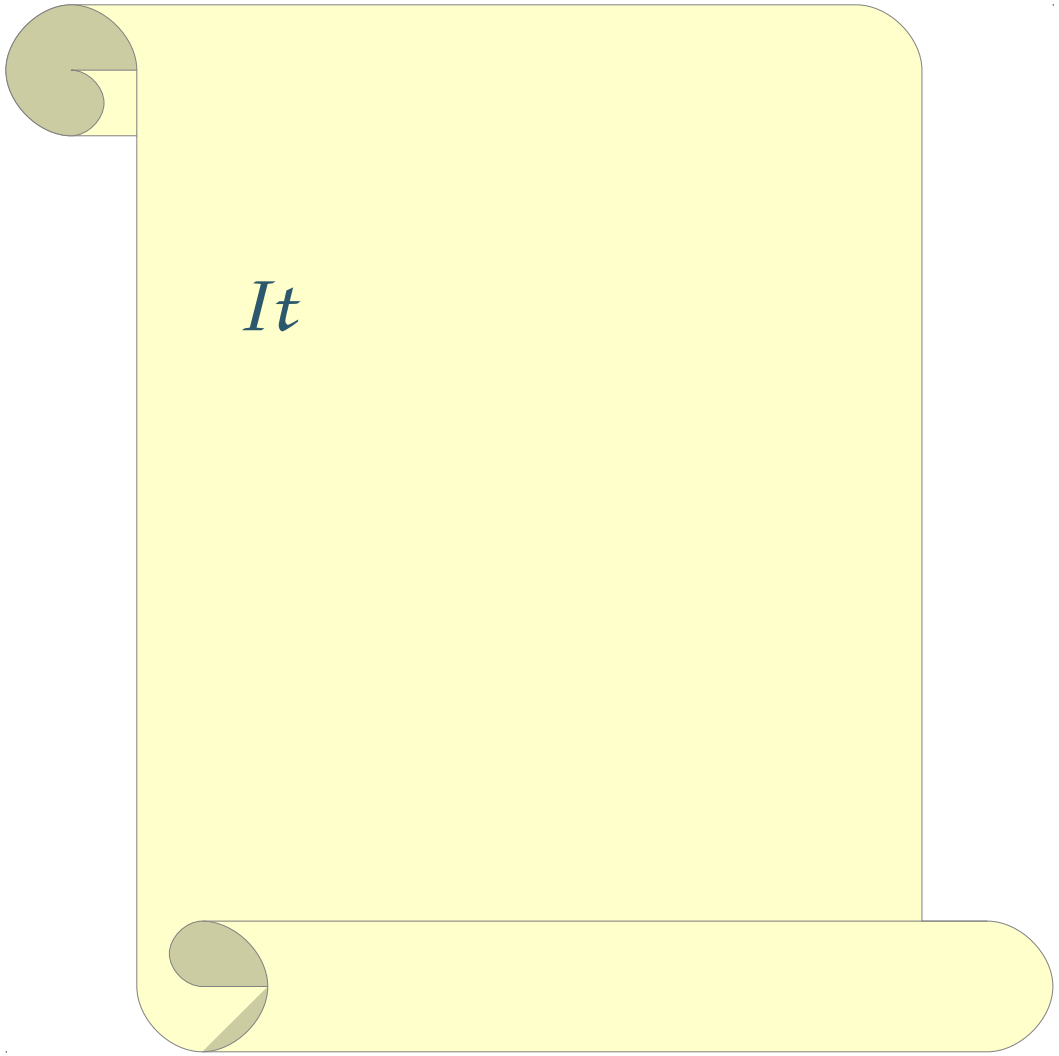

```
<surface n="1r" type="page">
  <label>1r</label>
  <zone type="column">
    <line>
      <hi rend="color:red">
        <w><hi rend="large">D</hi>er</w>
        <w>welhiſche</w>
        <w>gaſt</w>
        <pc>·</pc>
      </hi>
    </line>

    <line>
      <w>
        <hi rend="initial color:red">S</hi>
        <hi rend="decor-color:red">WER</hi>
      </w>
      <w>gern</w> <w>lıſt</w>
      <w>gvtıv</w>
      <w>m<rhyme n="2">êre</rhyme></w>
      <pc>·</pc>
```

An example

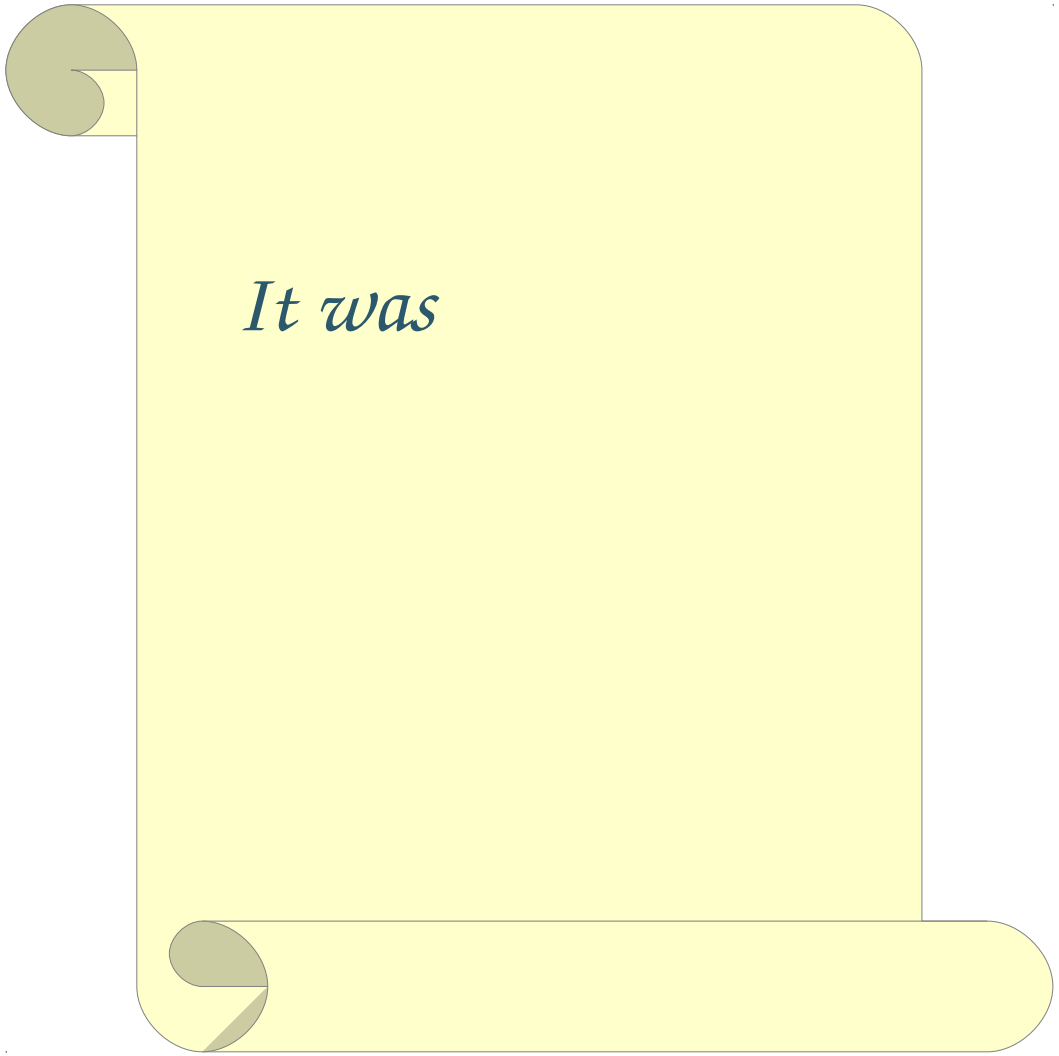



```
<page>
  <l>It was a
      <del>cold</del></l>
  <l>dark and</l>
  <l>stormy night</l>

  <date x="..." y="..."
        dt="1969-12-08">
      08/12/69
  </date>
</page>
```



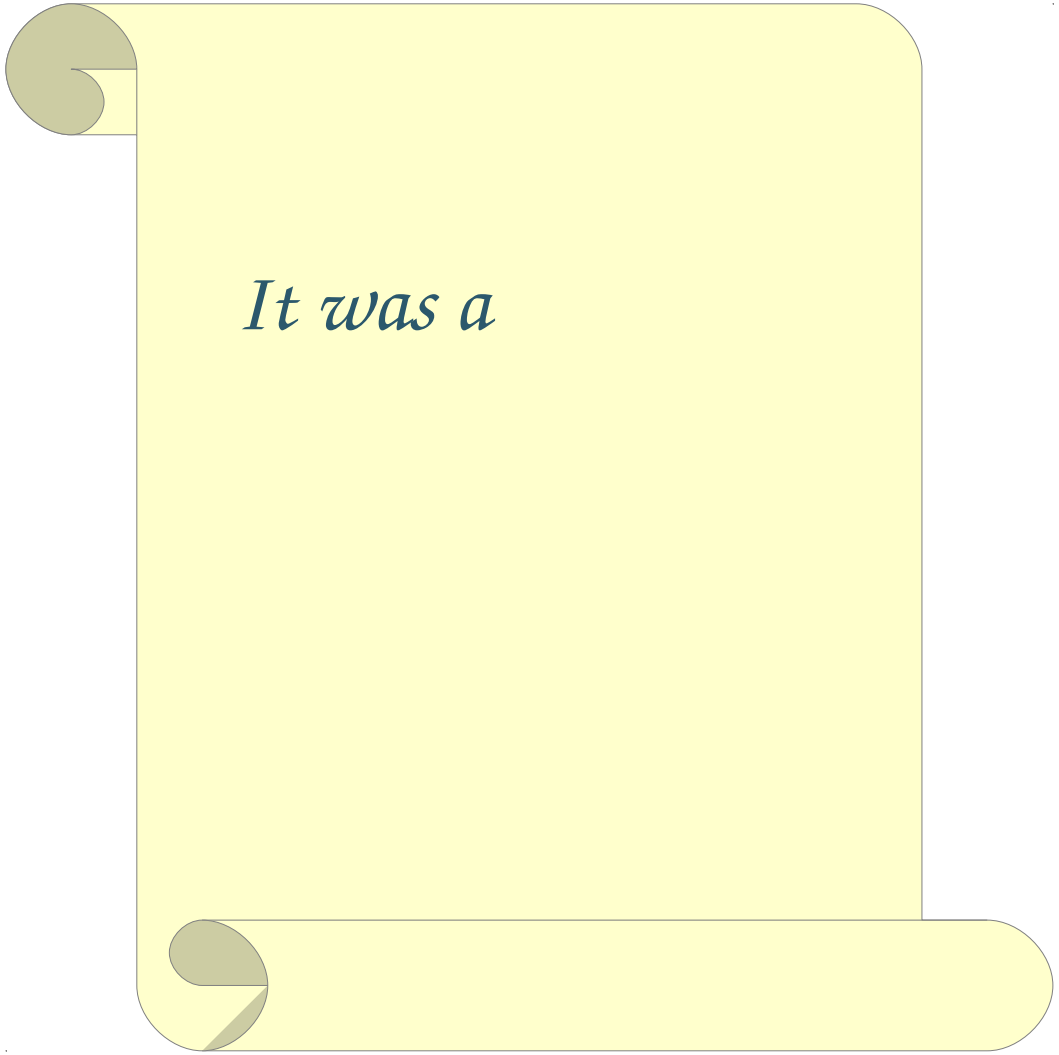


ADD



ADD

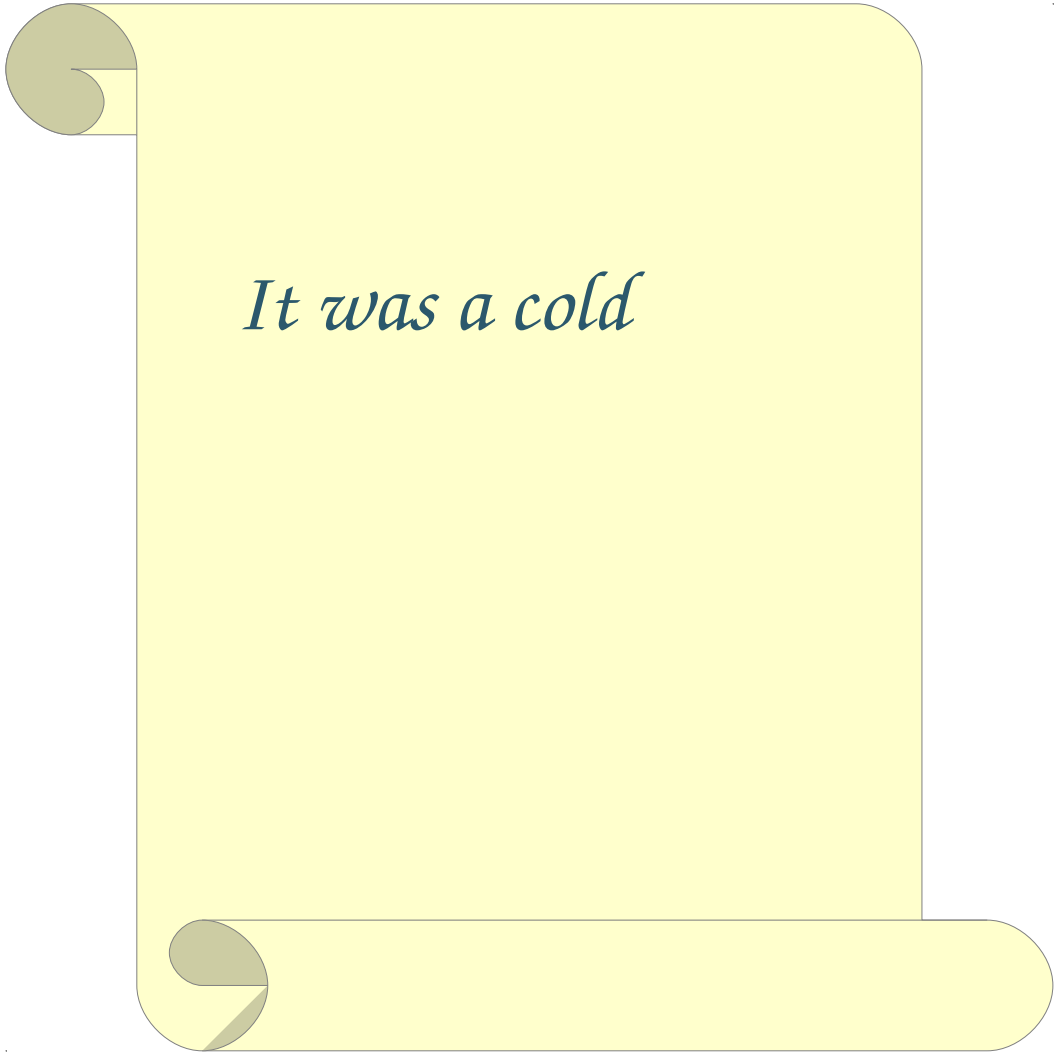
ADD



ADD

ADD

ADD

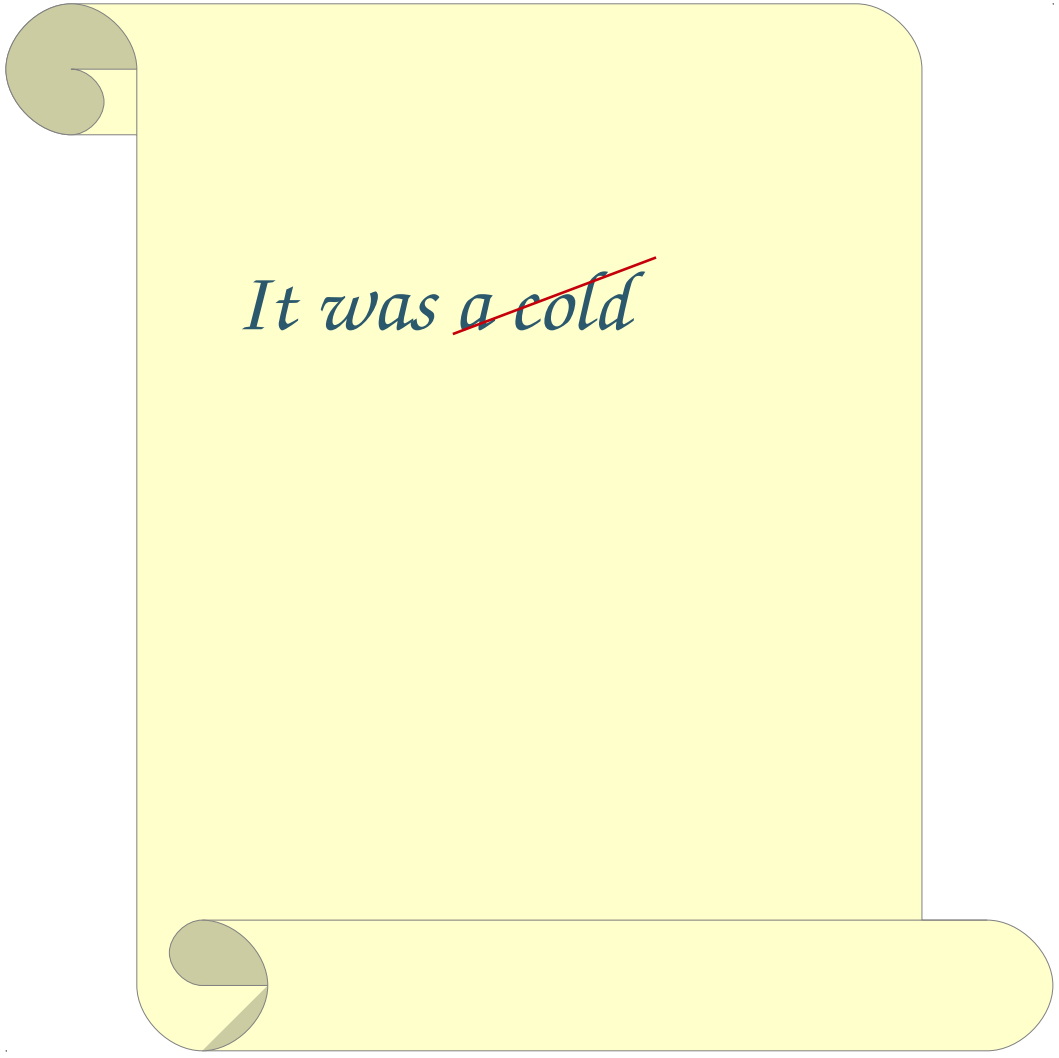


ADD

ADD

ADD

ADD



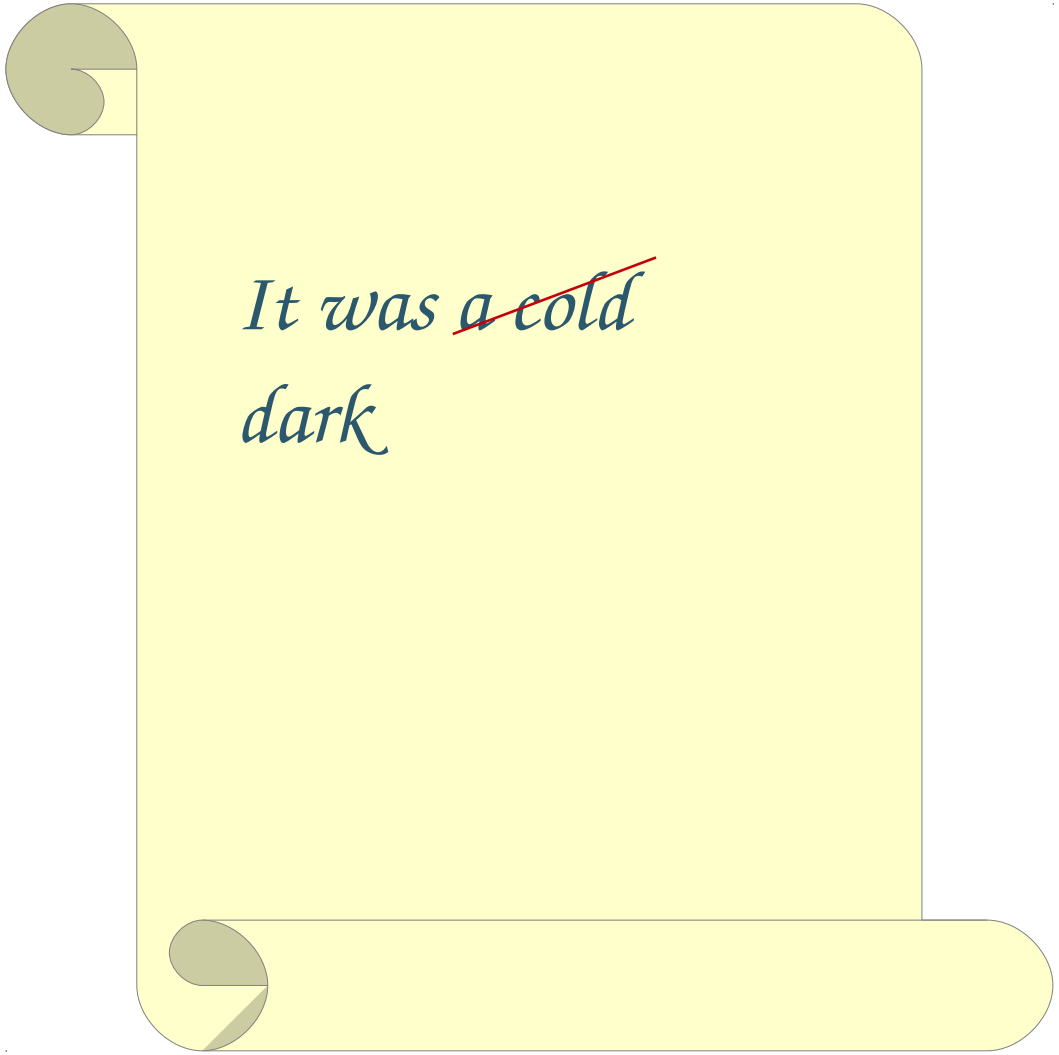
ADD

ADD

ADD

ADD

DEL



ADD

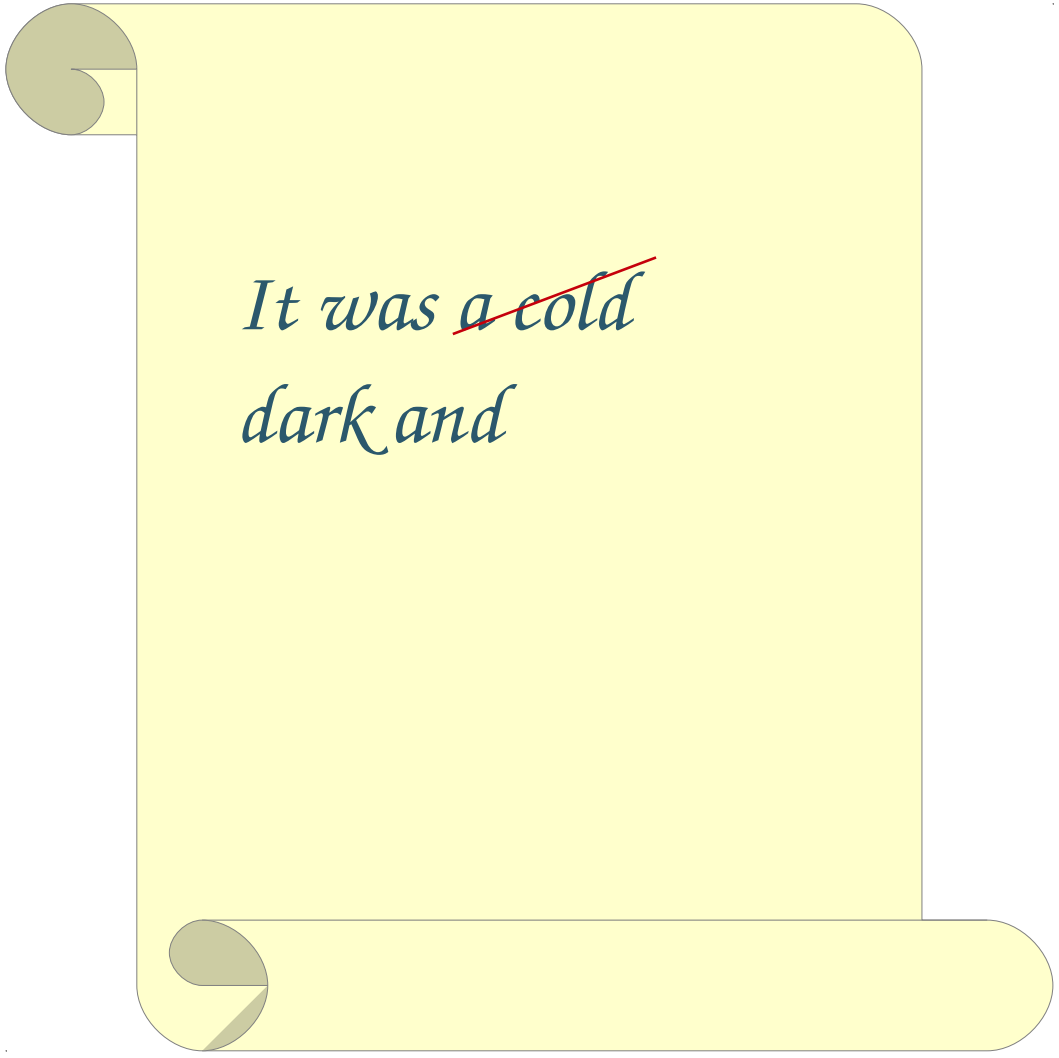
ADD

ADD

ADD

DEL

ADD



ADD

ADD

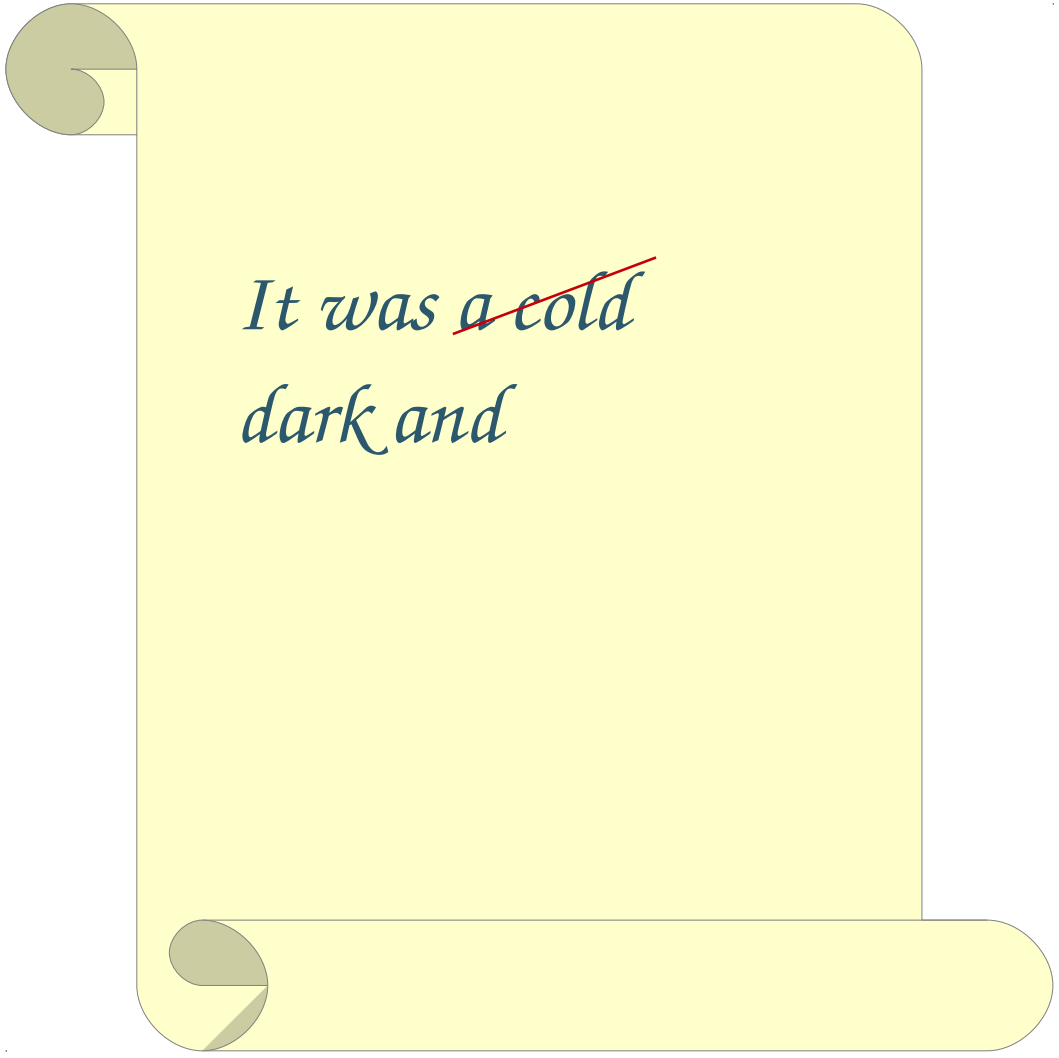
ADD

ADD

DEL

ADD

ADD



ADD(“It”, x, y)

ADD(“was”, …)

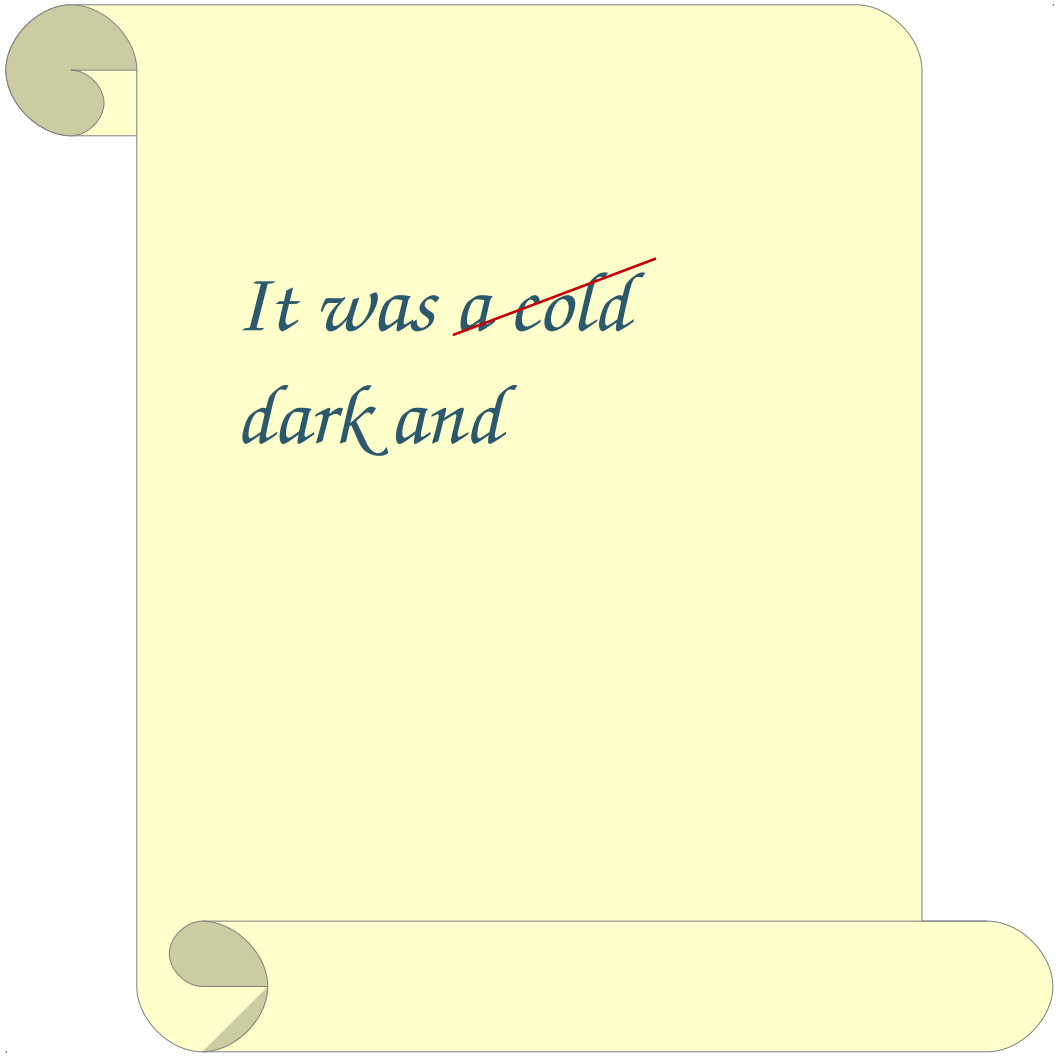
ADD(“a”, …)

ADD(“cold”, …)

DEL(“cold”, …)

ADD(“dark”, …)

ADD(“and”, …)



ADD(“It”, x, y)

ADD(“was”, …)

ADD(“a”, …)

ADD(“cold”, …)

DEL(“cold”, …)

ADD(“dark”, …)

ADD(“and”, …)

“Delta”

ADD("It", x, y)

ADD("was", …)

ADD("a", …)

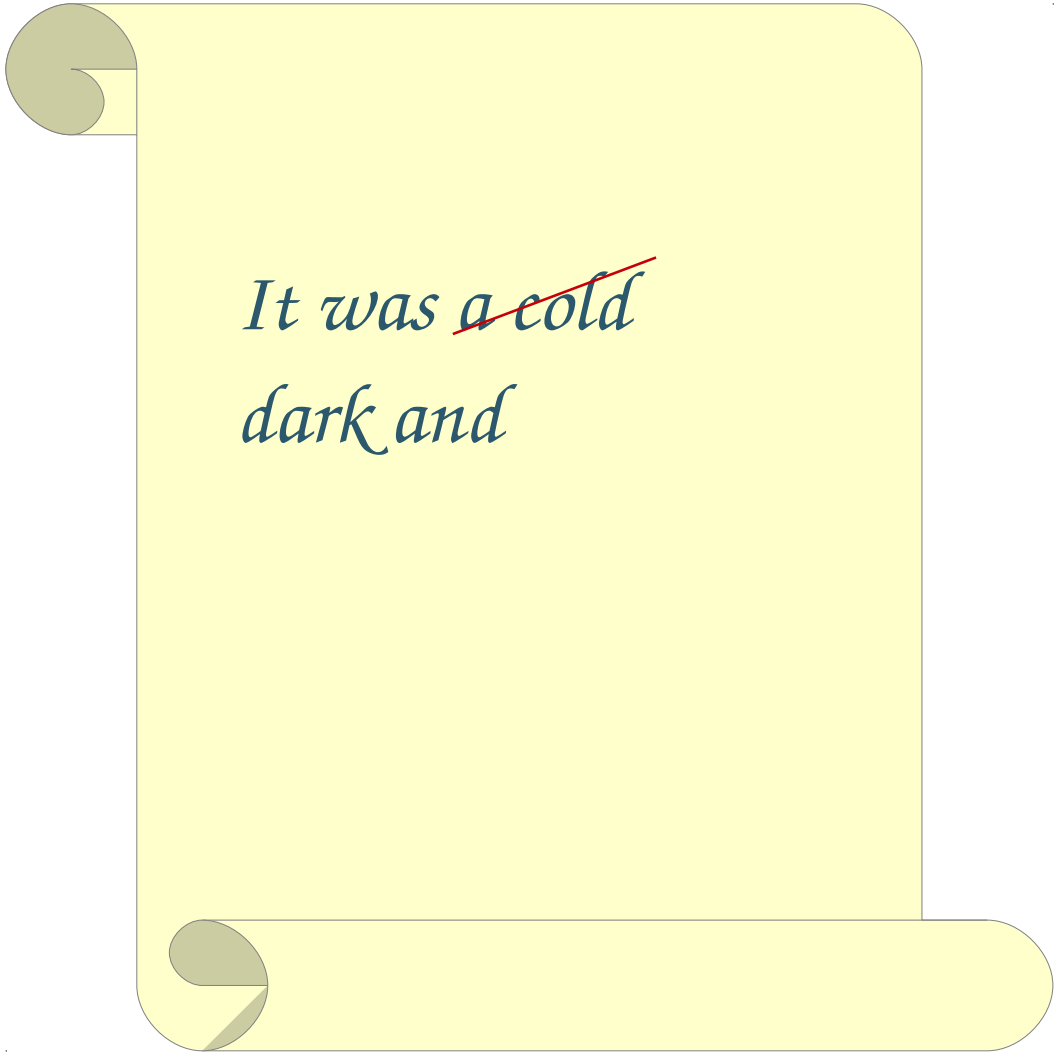
ADD("cold", …)

DEL("cold", …)

ADD("dark", …)

ADD("and", …)

"Change"



ADD(“It”, x, y)

ADD(“was”, …)

ADD(“a”, …)

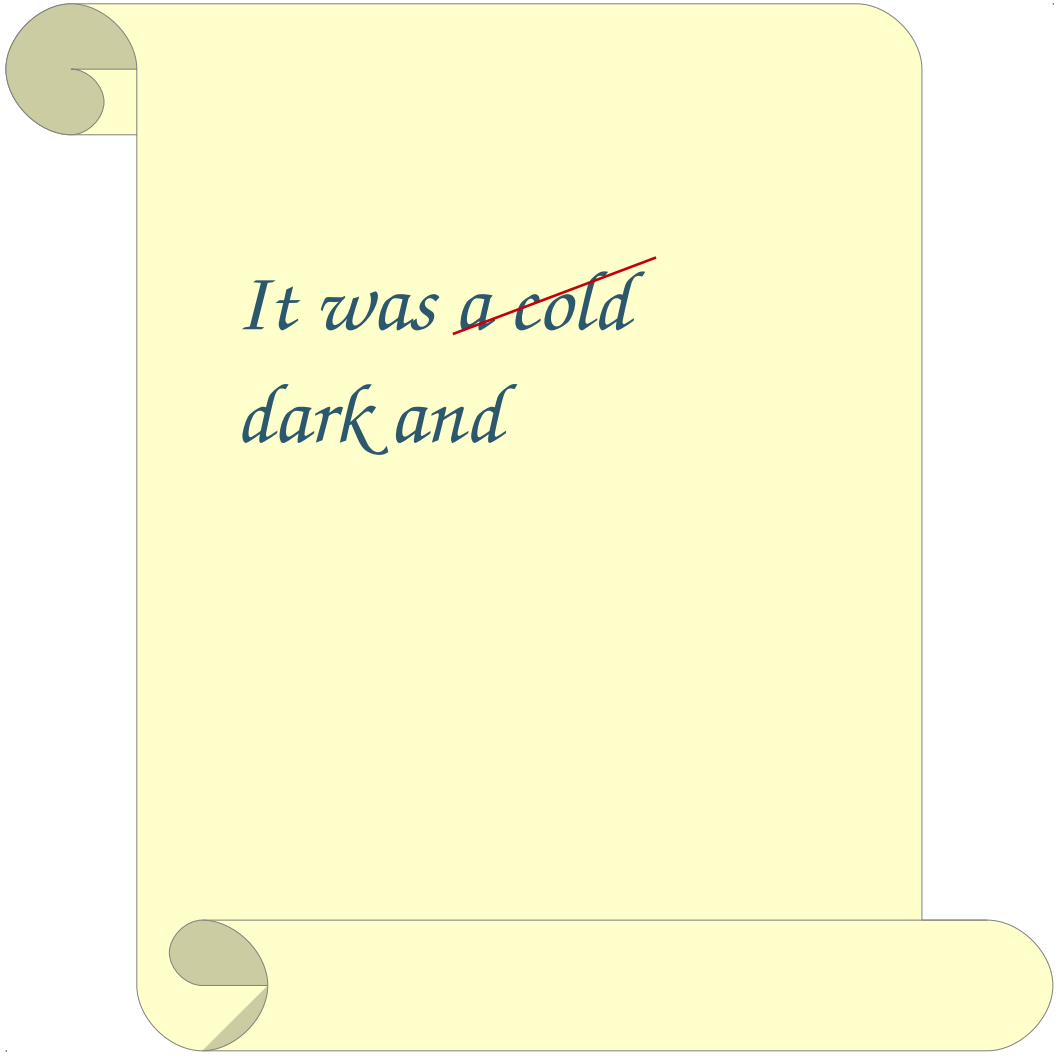
ADD(“cold”, …)

DEL(“cold”, …)

ADD(“dark”, …)

ADD(“and”, …)

“Operation”



“Scholarship”

ADD(“It”, x, y)		ADD(“It”, x, y)
ADD(“was”, …)		ADD(“was”, …)
ADD(“a”, …)		ADD(“a”, …)
ADD(“cold”, …)		ADD(“cold”, …)
DEL(“cold”, …)	“?”	ADD(“dark”, …)
ADD(“dark”, …)		DEL(“cold”, …)
ADD(“and”, …)		ADD(“and”, …)

Wuppertal 2020-02-06

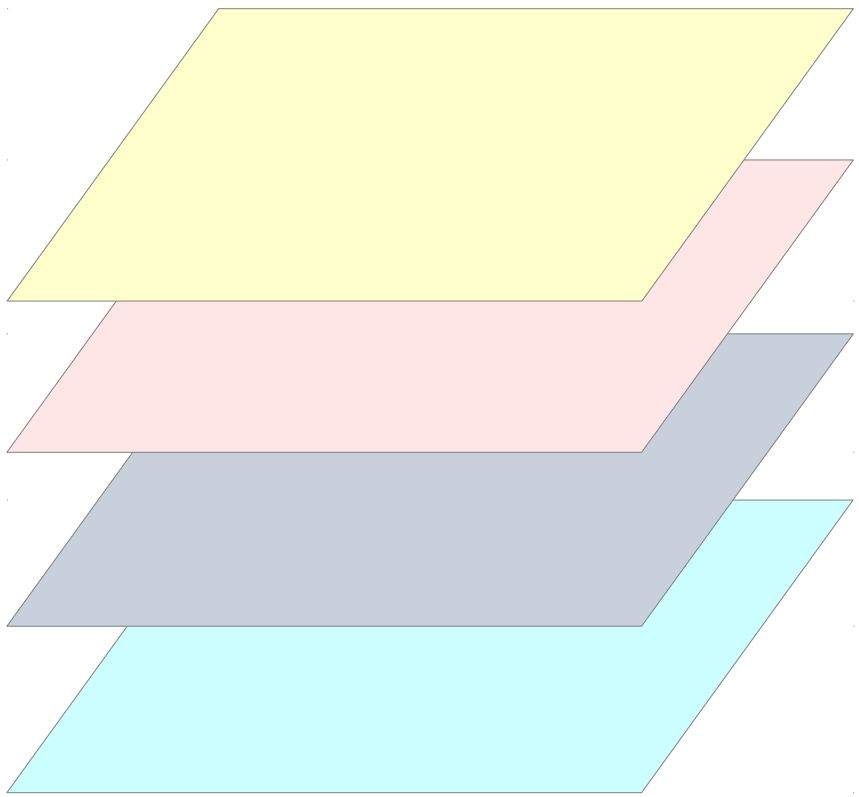
Ich meinte, sie sagte, er schrieb, dass er dachte...

Transcriptions as recordings of contextualized document changes

Gioele Barabucci
Institute of design

NTNU | Norwegian University of Science and Technology

Modification of what?



Text

Letters

Blobs of ink

Atoms

What is a digital document?

series of arguments

series of paragraphs

pieces of information

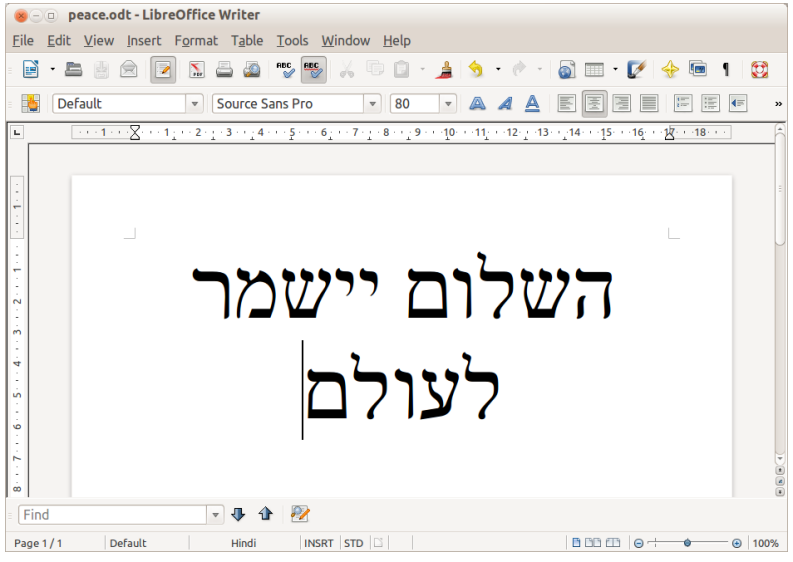
series of Unicode codepoints

series of bits

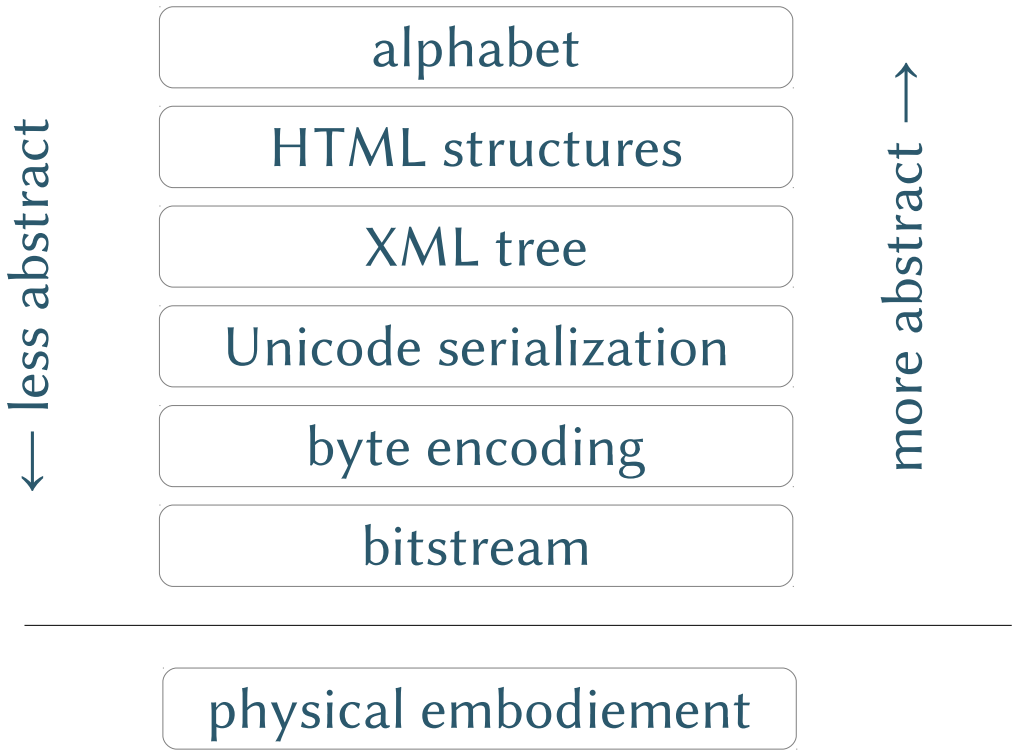
electrically charged surfaces

tree of XML nodes

pixels



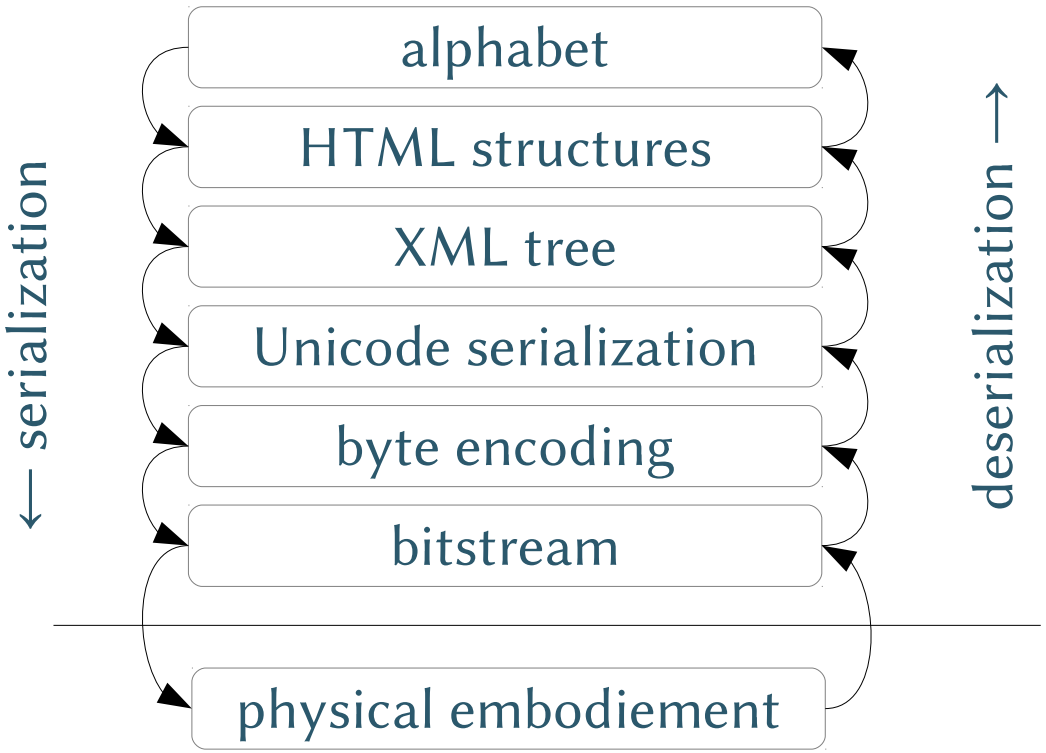
The CMV+P document model



- Content
- Model
  - atoms
  - structures
  - compositional rules
- Variants
  - possible alternative encodings

Gioele Barabucci. *The CMV+P document model, linear version*. In: Versioning Cultural Objects. 2019. IDE

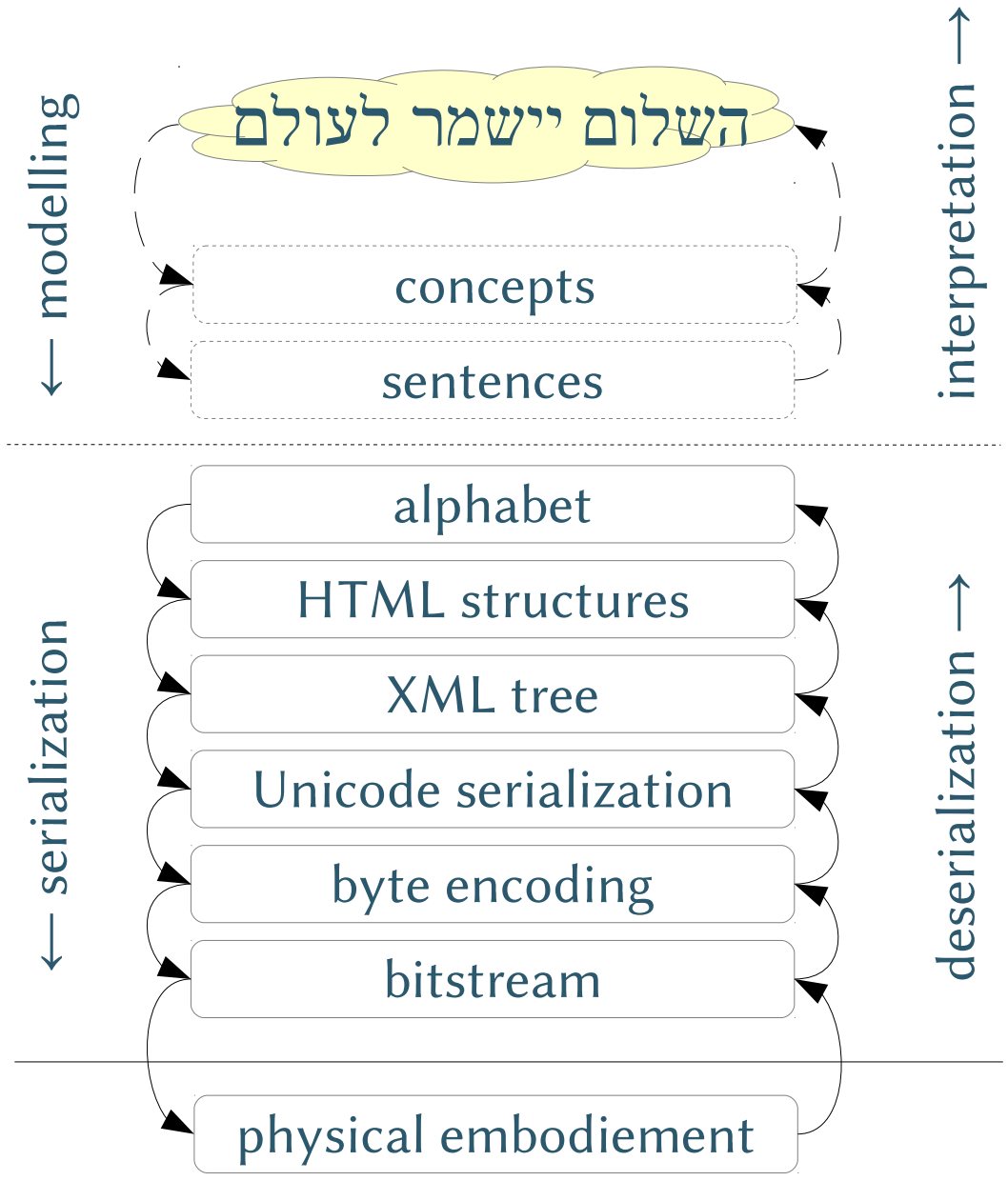
Transformation functions



Definition. A transformation function *trans* is a function that transforms the content C_a of an abstraction level L_a (created according to model M_a and variants V_a) into the content C_b of an abstraction level L_b, created according to the model M_b and the variants V_b.

$$trans : (C_a, M_a, V_a, M_b, V_b) \rightarrow C_b$$

Gioele Barabucci. *The CMV+P document model, linear version*. In: Versioning Cultural Objects. 2019. IDE



Stairway to abstraction

Point where the transformation functions stop being describable in mathematical terms.

Wuppertal 2020-02-06

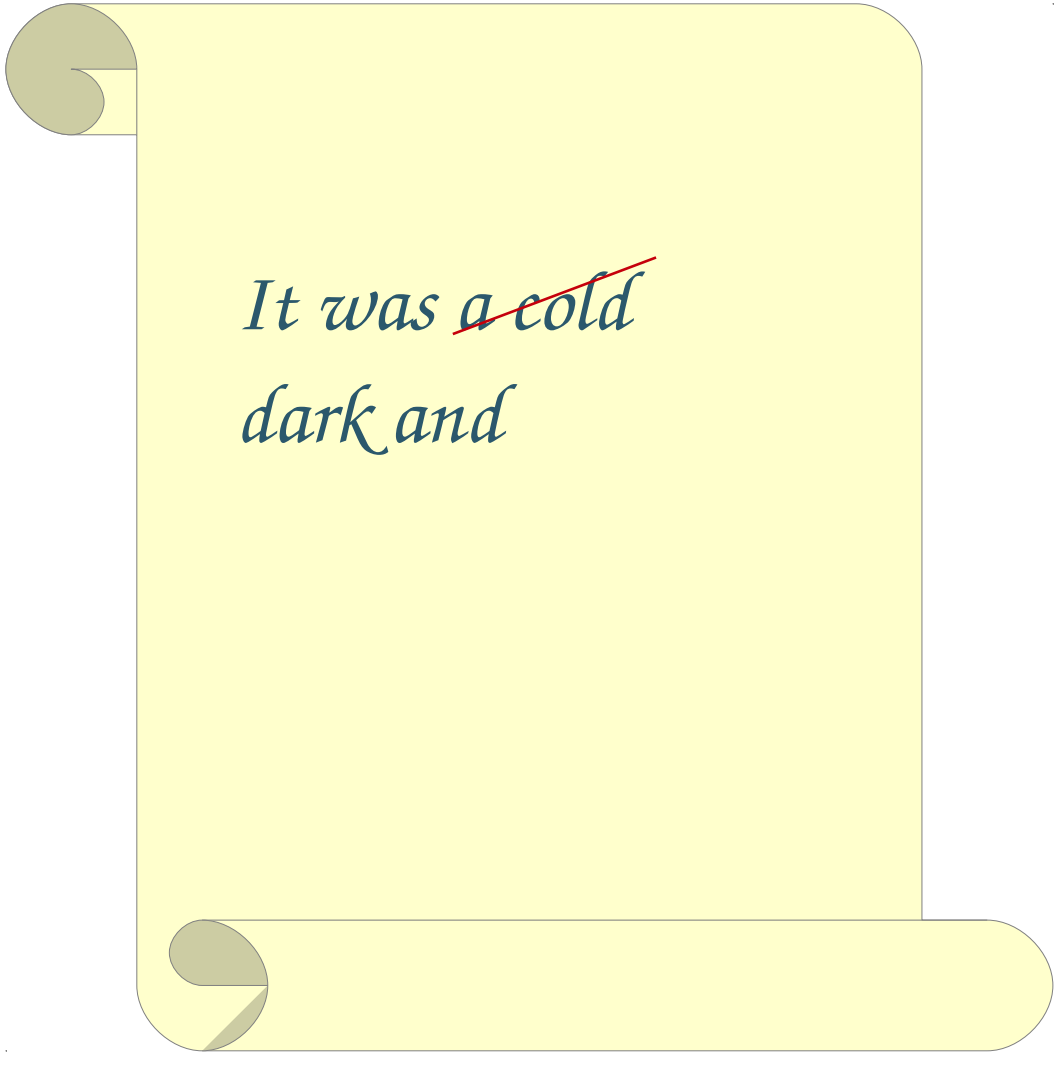
Ich meinte, sie sagte, er schrieb, dass er dachte...

Transcriptions as recordings of contextualized document changes

Gioele Barabucci
Institute of design

NTNU | Norwegian University of Science and Technology

CMV+P and deltas



ADD(“It”, x, y)

ADD(“was”, …)

ADD(“a”, …)

ADD(“cold”, …)

DEL(“cold”, …)

ADD(“dark”, …)

ADD(“and”, …)

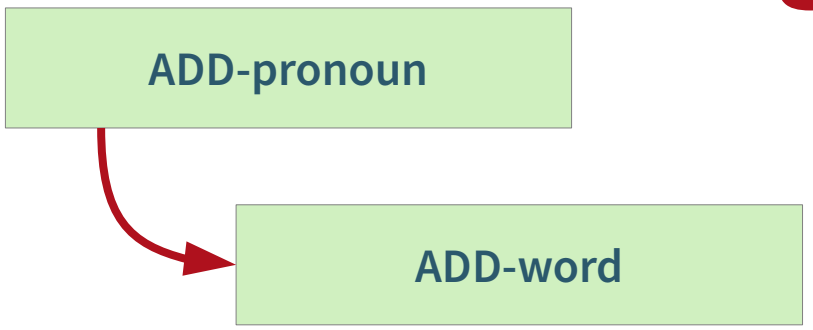
- ADD-pronoun
- ADD-word
- ADD-letter
- ADD-stroke
- ADD-ink-drops

Changes at different levels of abstraction

ADD-pronoun

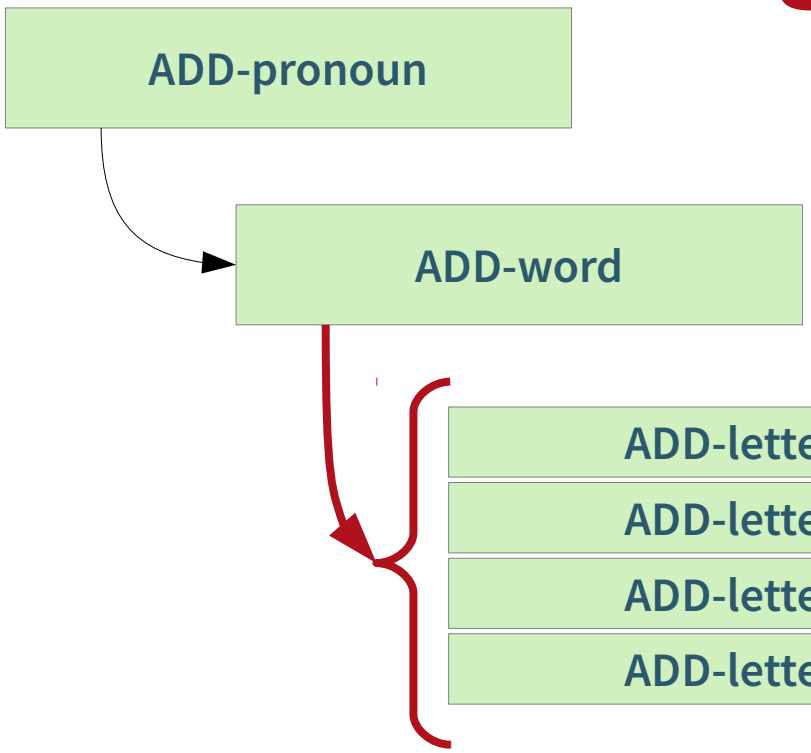
Structured (meaningful) changes

Structured (meaningful) changes



+ `word` is one of {I, you, he, she, it, we, they}

Structured (meaningful) changes

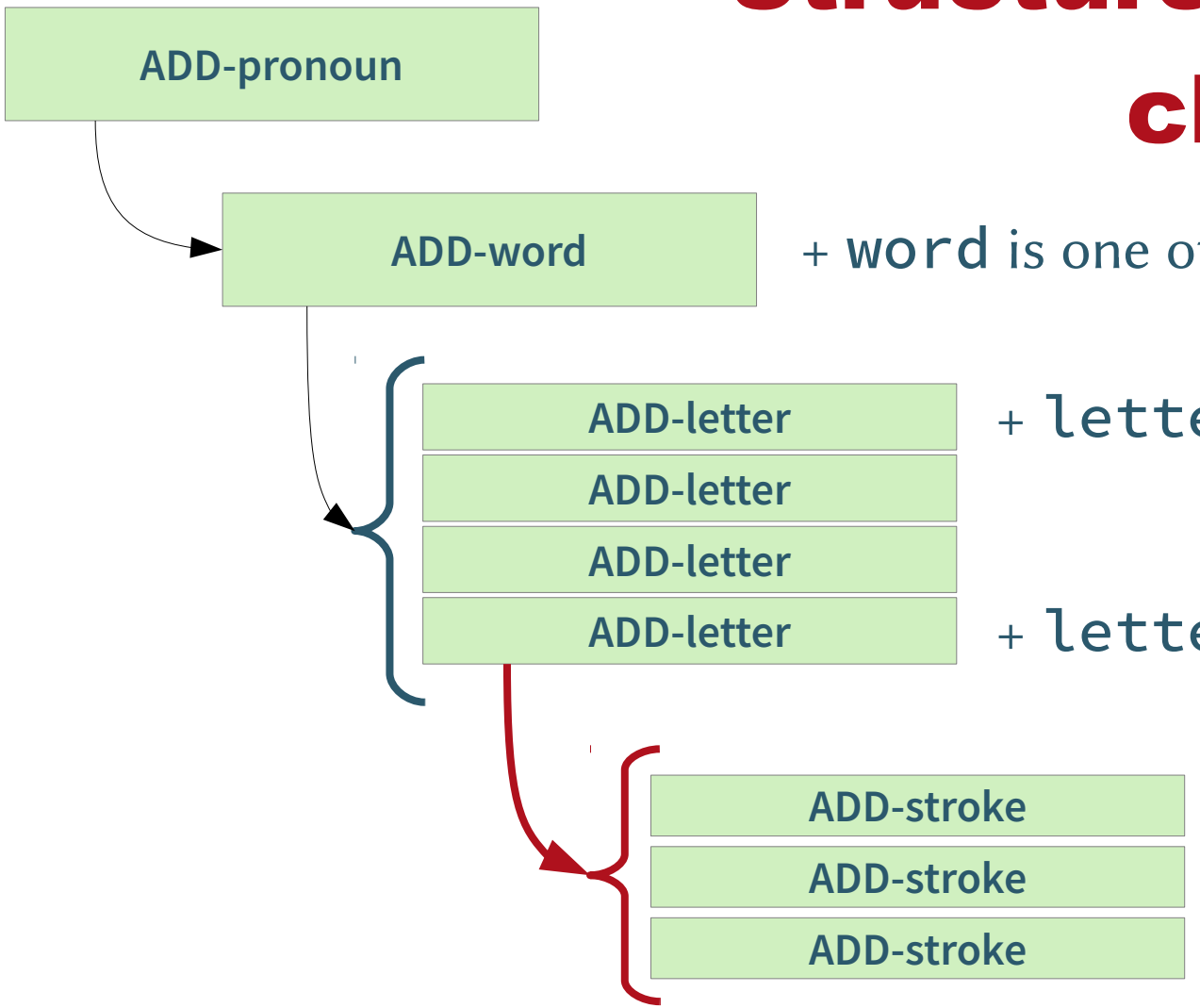


+ `word` is one of {I, you, he, she, it, we, they}

+ `letter` is not “ “

+ `letter` is not “ “ or **PUNCTUATION**

Structured (meaningful) changes



+ `word` is one of {I, you, he, she, it, we, they}

+ `letter` is not “ “

+ `letter` is not “ “ or **PUNCTUATION**

Formalization

$$
\begin{aligned}
&\exists c_1, c_2 \mid \\
&\quad c_1.time < c_2.time \wedge \\
&\quad c_1.op = DEL \wedge c_2.op = ADD \wedge \\
&\quad c_1.idx = c_2.idx \\
&\Rightarrow \\
&\quad c_n = REPLACE\left(c_1.data, c_2.data, c_1.idx\right)
\end{aligned}
$$

Facts or opinions?

~~Facts or opinions?~~

There are only opinions!

Raw data

Castner Wolphgangus è un de li studenti tedeschi

From raw data to structured data

Person {

Id = 0918

Name = Castner Wolphgangus

Occupation = Student

Origin = Germany

}

Data can be temporally bound

Castner Wolphgangus è un de li studenti tedeschi

Annali 1570-1571

Data can be imprecise

Cas__er WoIphga è un de li studenti 'deschi

Data can be incoherent

Castner Wolphgangus è un de li studenti tedeschi

Marcus Bononii et Wolfgangus Kastner amici sunt

Possible solution: factoids + PROSO

- Just be explicit about authorship, assumptions, sources, etc.
- PROSO factoids are inspired by Bradley and Short, 2005: “The evolving field of new-style prosoprography”.

Author Gioele Barabucci states that
Source Annali 1570-1571 states that
Wofgang Castner was a student since 1571

Wuppertal 2020-02-06

Ich meinte, sie sagte, er schrieb, dass er dachte...

Transcriptions as recordings of contextualized document changes

Gioele Barabucci
Institute of design

NTNU | Norwegian University of Science and Technology

Do you want hyper-diplomacy?

- Documents did non fall out of the blue:
  - Time dimension is important.
  - Document as series of incremental modification.
  - Transcription as series of incremental changes.
- Physical and digital documents exist at many different levels of abstraction:
  - Do not throw everything in one pot.
  - Explicit recording of which level one is referring to.
  - Composition of higher-level assertions on top of lower-level assertions.
- Scholarship is made of opinions based on other opinions:
  - Explicit recording of context, authorship, provenance with every assertion

Ich meinte, sie sagte,
er schrieb, dass er dachte…

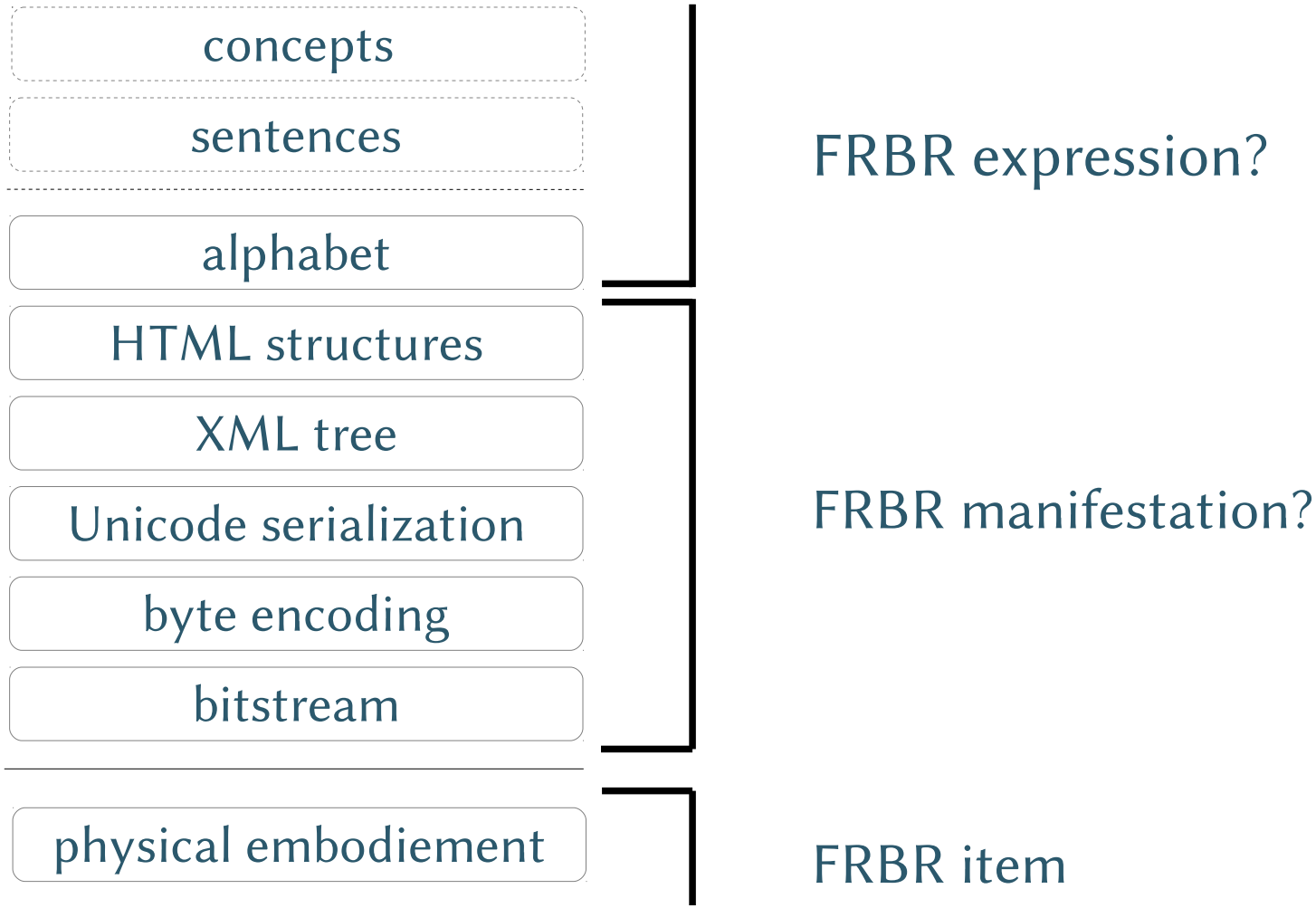
Transcriptions as recordings of
contextualized document changes

Thank you!

Gioele Barabucci
Institute of design

NTNU | Norwegian University of
Science and Technology

(Tiny) Bibliography

- Deltas / Meaningful changes
  - Manuscript annotations as deltas: first steps. Barabucci. 2016.
- Levels of abstraction / CMV+P
  - The CMV+P Document Model, Linear Version. Barabucci. 2019.
- Context / PROSO
  - "The evolving field of new-style prosoprography". Bradley and Short. 2005.
  - PROSO: prosopographic records. Barabucci and Zingoni. 2013.

CMV+P vs FRBR Group 1 entities