

Datenformate in den Digital Humanities (E01)



Prof. Dr. Christof Schöch

Modul Auszeichnungssprachen
MSc. Digital Humanities, Universität Trier





Überblick

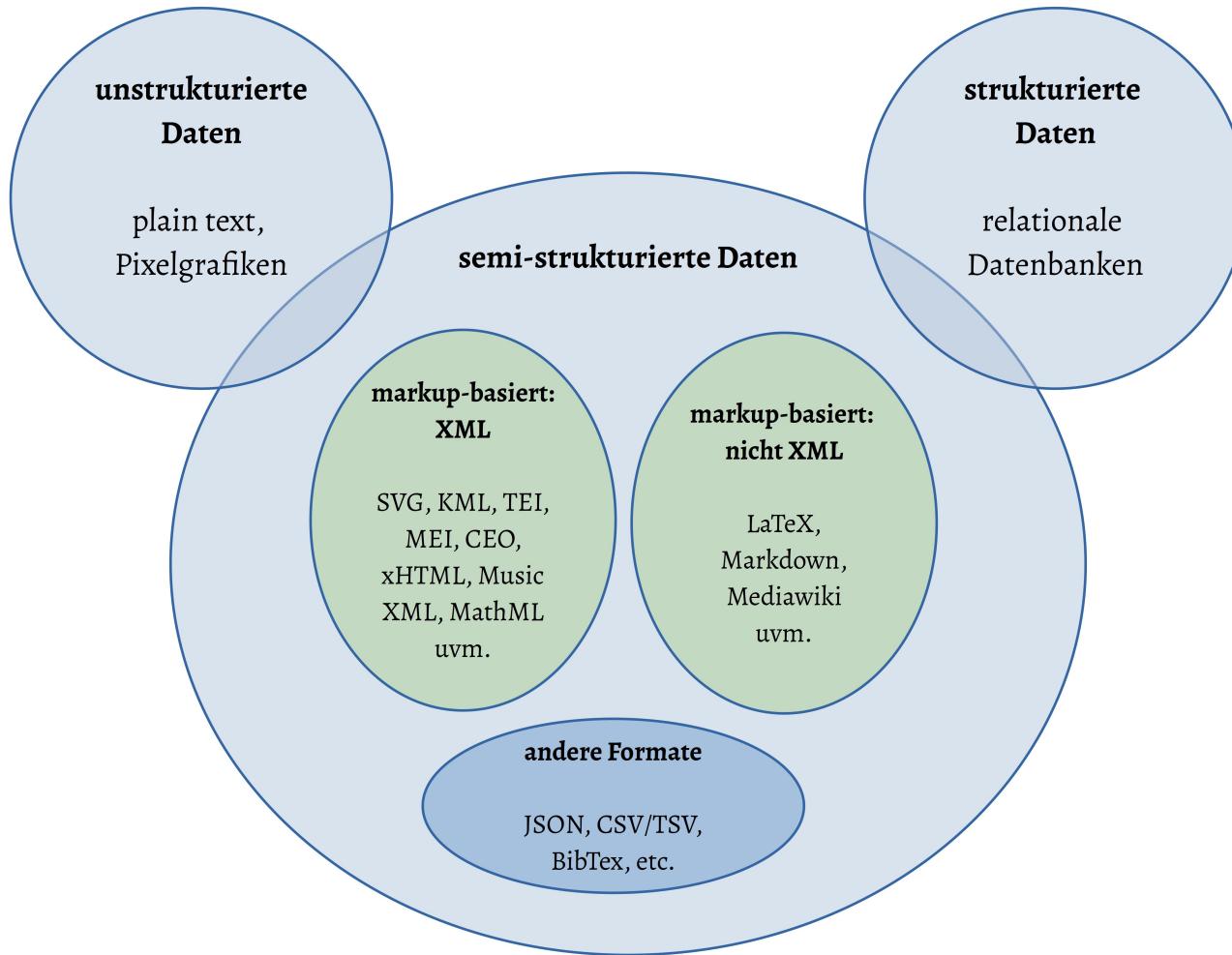
1. Datenformate im Überblick
2. Datenformate im Einzelnen

(1) Datenformate im Überblick

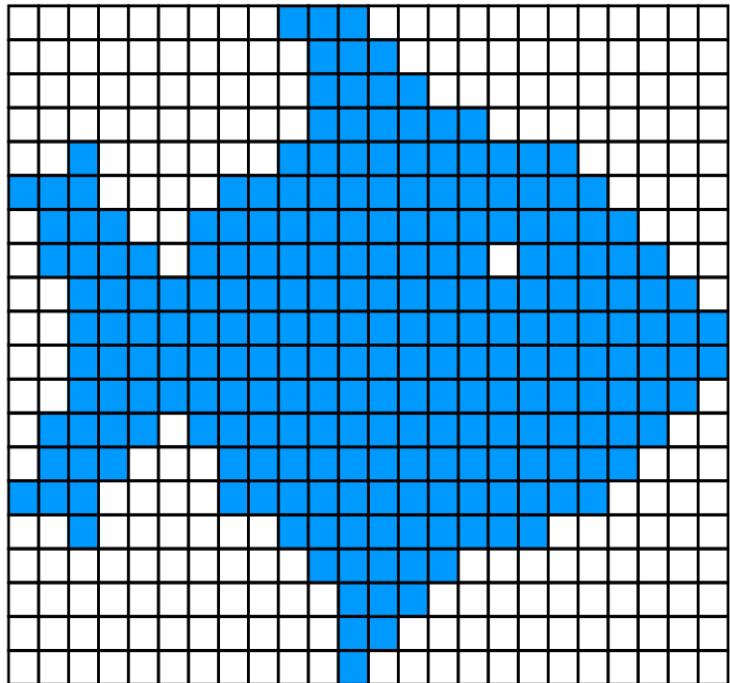
Schichten der Datenrepräsentation

1. Untersuchungsgegenstand: realweltlicher Gegenstand
2. Modellierte Instanz: digitale Repräsentation des Gegenstands
3. Datenformate: strukturieren die Informationen
4. Zeichenfolge: unstrukturierte Abfolge von Zeichen
5. Zeichenkodierung: bspw. Unicode/UTF-8: Code für jedes Zeichen
6. Binärcode: Repräsentation der Zeichencodes im Speicher

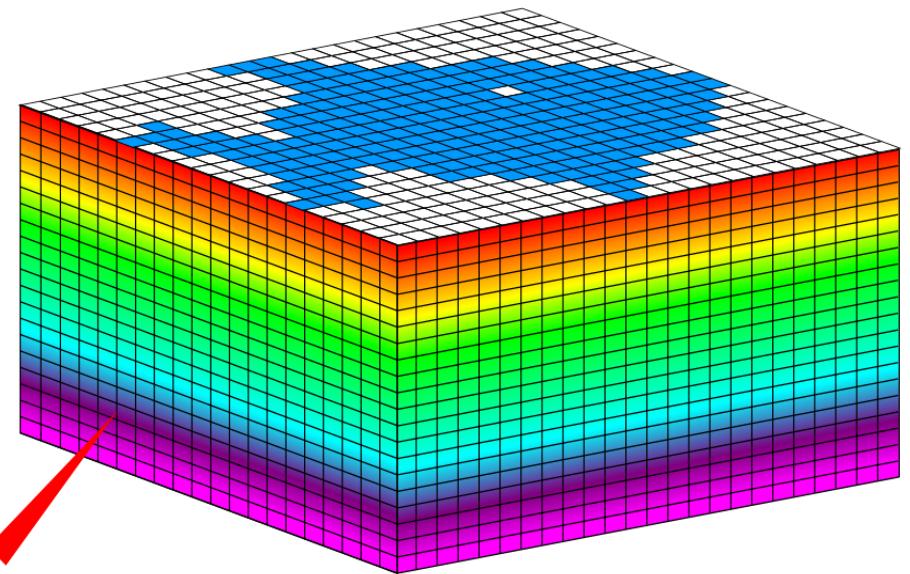
Datenformate in den DH



Pixelgrafik



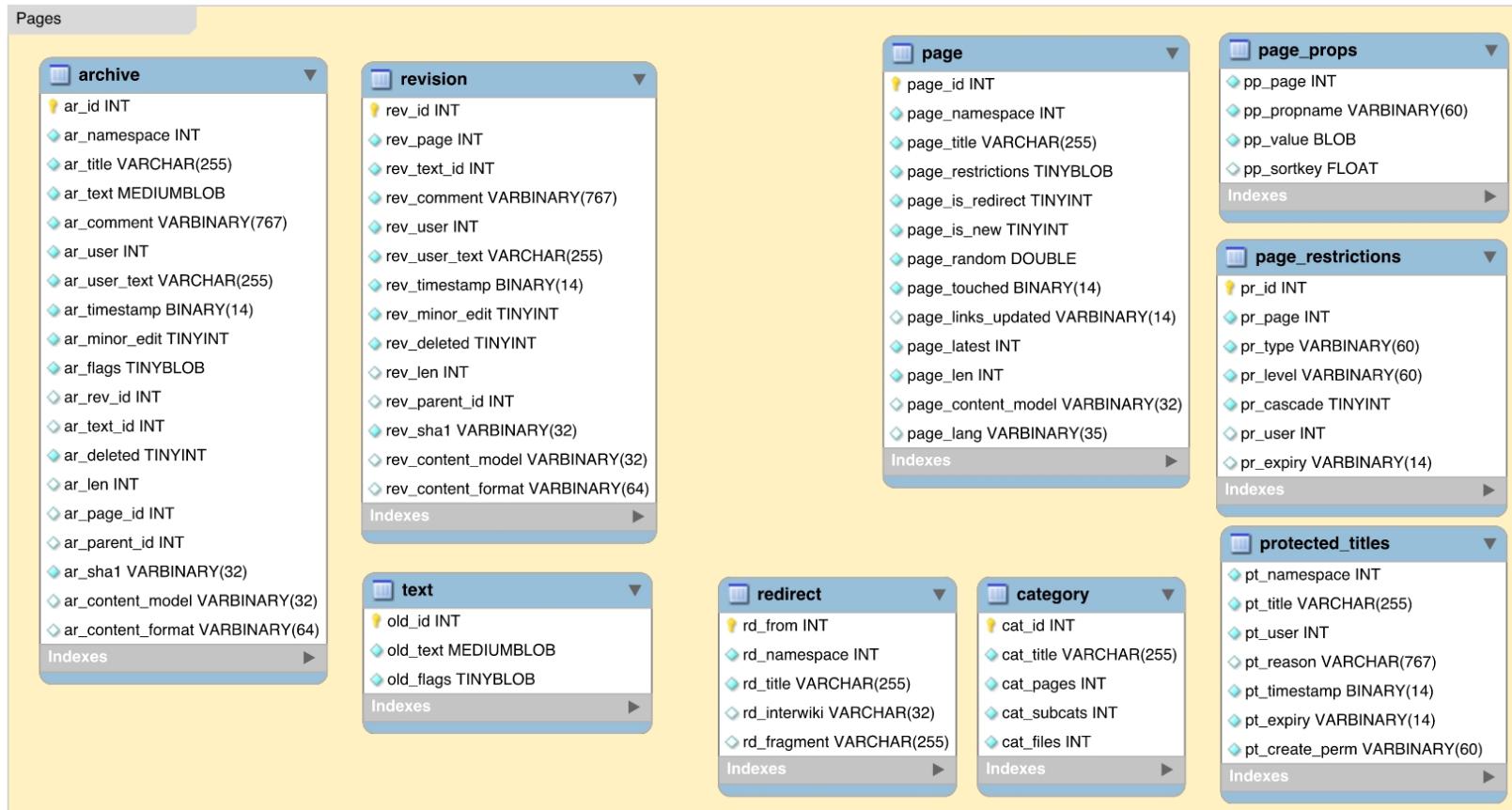
Punktdichte



Farbtiefe

(Quelle: Wikipedia: "Punktdichte+Farbtiefe", Nutzer Friedrich Graf, Lizenz CC BY-SA.)

Datenbanken



(Quelle: [Wikipedia Database Schema](#), Nutzer Timo Tijhof, CC BY.)

Semi-strukturierte Formate

Semi-strukturierte Formate

- Nicht Markup-basiert
 - **CSV/TSV**: Comma/Tab-Separated Values
 - **JSON**: JavaScript Object Notation (tweet)
 - **BibLaTeX**

Semi-strukturierte Formate

- Nicht Markup-basiert
 - **CSV/TSV**: Comma/Tab-Separated Values
 - **JSON**: JavaScript Object Notation (tweet)
 - **BibLaTeX**
- Markup-basiert (XML)
 - **HTML**: HyperText Markup Language
 - **XML**: eXtensible Markup Language
 - **KML**: Keyhole Markup Langauge
 - **TEI**: Text Encoding Initiative
 - **SVG**: Scalable Vector Graphics (live)

Semi-strukturierte Formate

- Nicht Markup-basiert
 - **CSV/TSV**: Comma/Tab-Separated Values
 - **JSON**: JavaScript Object Notation (tweet)
 - **BibLaTeX**
- Markup-basiert (XML)
 - **HTML**: HyperText Markup Language
 - **XML**: eXtensible Markup Language
 - **KML**: Keyhole Markup Langauge
 - **TEI**: Text Encoding Initiative
 - **SVG**: Scalable Vector Graphics (live)
- Markup-basiert (nicht XML)
 - **LaTeX**
 - **MediaWiki**
 - **Markdown**

(2) Dateiformate im Einzelnen

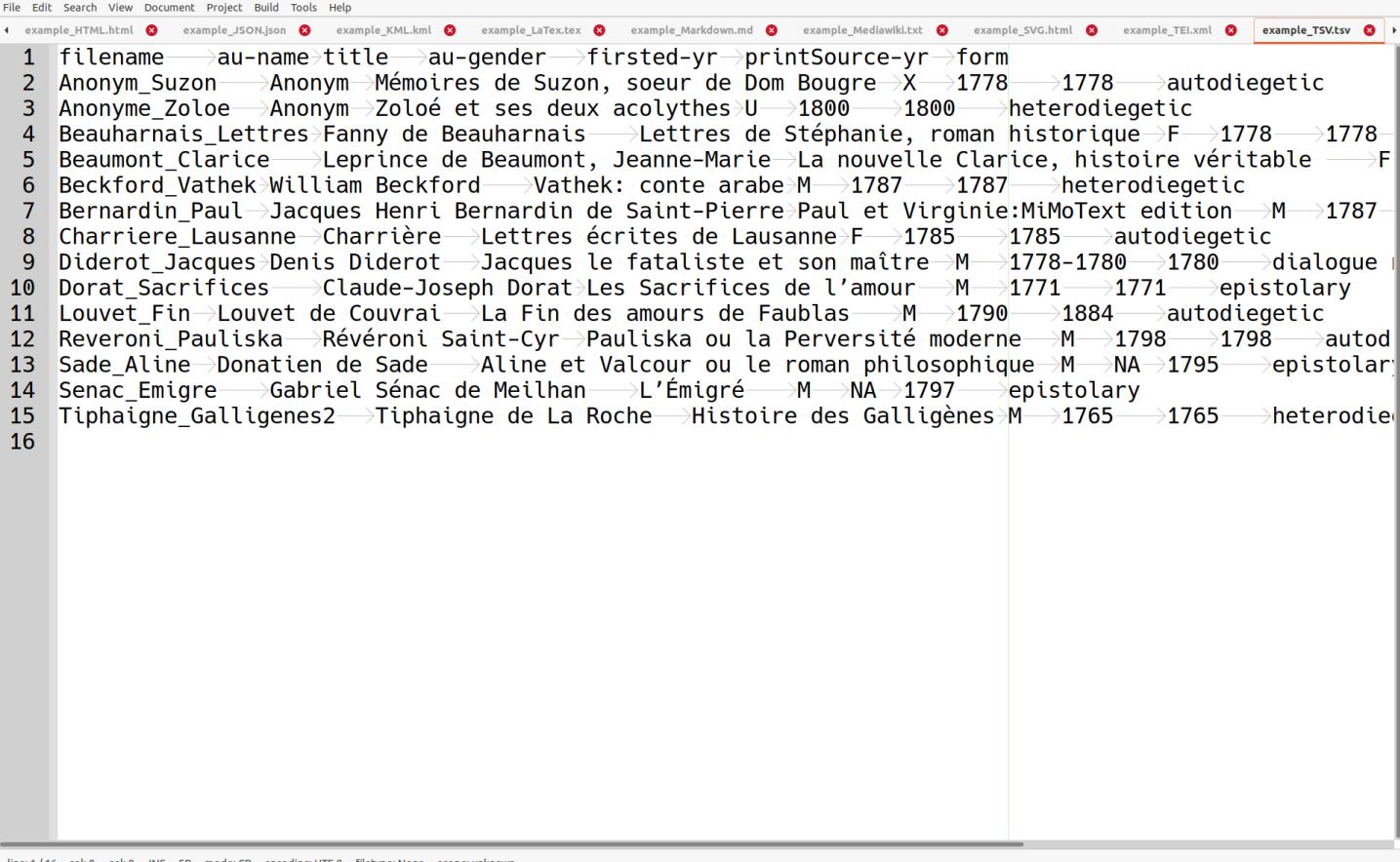
Was macht Datenformate geeignet?

- Weit verbreitet und standardisiert
- Nicht proprietär
- Offen dokumentiert
- Verlustfreie oder keine Kompression
- Einfach dekodierbar oder unmittelbar lesbar

(Nach: "IT-Empfehlungen. Für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften", IANUS, Version 1.0.1.0, 2017.
<https://www.ianus-fdz.de/it-empfehlungen/.>)

Nicht Markup-basiert

CSV/TSV



The screenshot shows a text editor window with a menu bar (File, Edit, Search, View, Document, Project, Build, Tools, Help) and a tab bar at the top. The tab bar includes 'example_HTML.html', 'example_JSON.json', 'example_KML.kml', 'example_LaTex.tex', 'example_Markdown.md', 'example_Mediawiki.txt', 'example_SVG.html', 'example_TEI.xml', and 'example_TSV.tsv'. The 'example_TSV.tsv' tab is selected.

The main content area displays a CSV/TSV file with 16 entries, each starting with a number from 1 to 16. The columns represent fields such as author, title, gender, first published year, print source year, form, and genre.

	filename	au-name	title	au-gender	firsted-yr	printSource-yr	form	genre
1	Anonym_Suzon	Anonym	Mémoires de Suzon, soeur de Dom Bougre	X	1778	1778	autodiegetic	
2	Anonyme_Zoloe	Anonym	Zoloé et ses deux acolythes	U	1800	1800	heterodiegetic	
3	Beauharnais_Lettres	Fanny de Beauharnais	Lettres de Stéphanie, roman historique	F	1778	1778		
4	Beaumont_Clarice	Leprince de Beaumont, Jeanne-Marie	La nouvelle Clarice, histoire véritable	F				
5	Beckford_Vathek	William Beckford	Vathek: conte arabe	M	1787	1787	heterodiegetic	
6	Bernardin_Paul	Jacques Henri Bernardin de Saint-Pierre	Paul et Virginie: MiMoText edition	M	1787			
7	Charriere_Lausanne	Charrière	Lettres écrites de Lausanne	F	1785	1785	autodiegetic	
8	Diderot_Jacques	Denis Diderot	Jacques le fataliste et son maître	M	1778-1780	1780	dialogue	
9	Dorat_Sacrifices	Claude-Joseph Dorat	Les Sacrifices de l'amour	M	1771	1771	epistolary	
10	Louvet_Fin	Louvet de Couvrai	La Fin des amours de Faublas	M	1790	1884	autodiegetic	
11	Reveroni_Pauliska	Réveroni Saint-Cyr	Pauliska ou la Perversité moderne	M	1798	1798	autod	
12	Sade_Aline	Donatien de Sade	Aline et Valcour ou le roman philosophique	M	NA	1795	epistolary	
13	Senac_Emigre	Gabriel Sénac de Meilhan	L'Émigré	M	NA	1797	epistolary	
14	Tiphaigne_Galligenes2	Tiphaigne de La Roche	Histoire des Galligènes	M	1765	1765	heterodie	
15								
16								

At the bottom of the editor window, there is a status bar with the following information: line: 1 / 16, col: 0, sel: 0, INS, SP, mode: CR, encoding: UTF-8, filetype: None, scope: unknown.

CSV/TSV: Comma/Tab-Separated Values

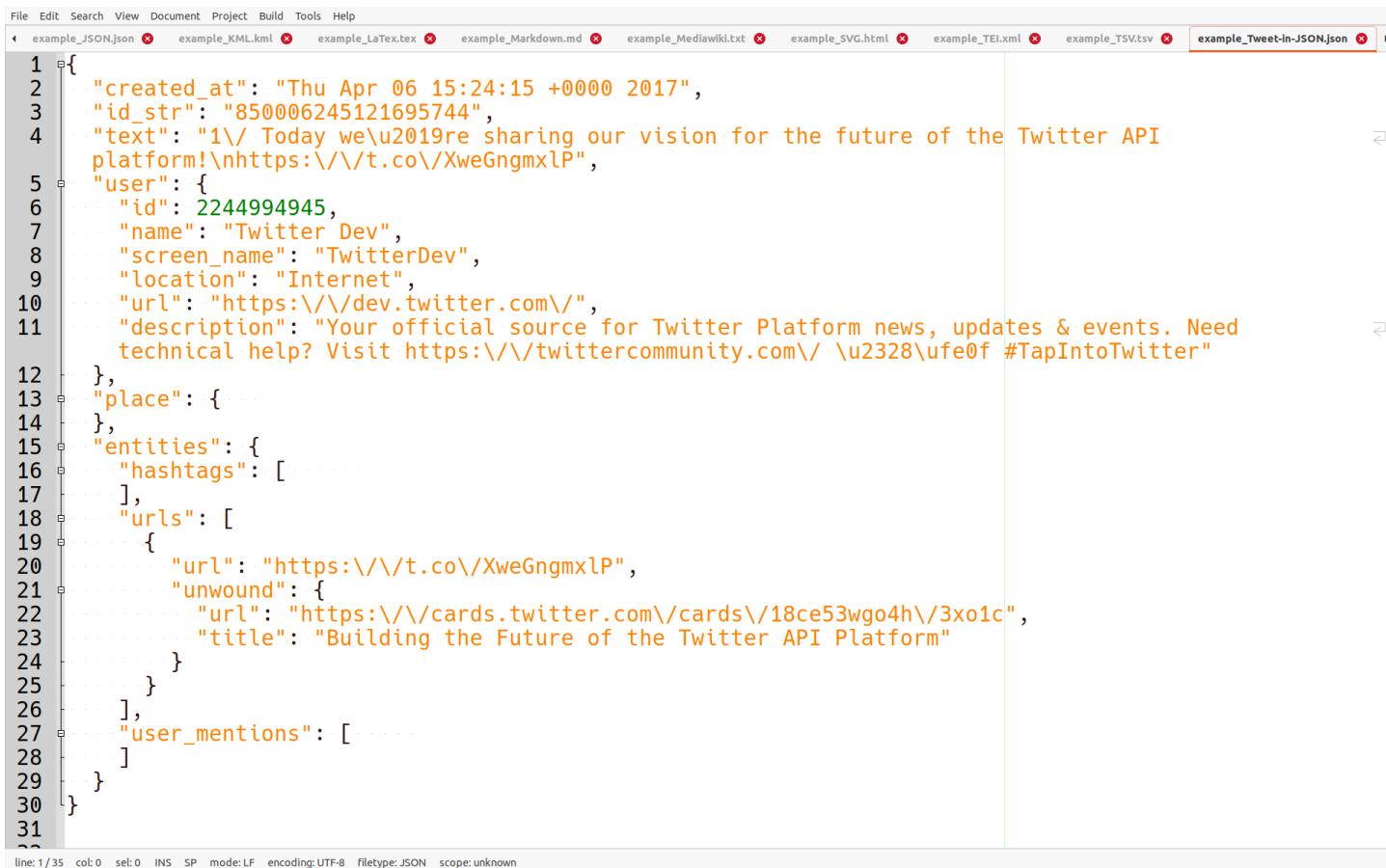
JSON

```
1 {
2   "links": {
3     "self": "https://www.rijksmuseum.nl/api/nl/collection/SK-C-5",
4     "web": "https://www.rijksmuseum.nl/nl/collectie/SK-C-5"
5   },
6   "id": "nl-SK-C-5",
7   "objectNumber": "SK-C-5",
8   "title": "Schutters van wijk II onder leiding van kapitein Frans Banninck Cocq,  
bekend als de 'Nachtwacht'",
9   "hasImage": true,
10  "principalOrFirstMaker": "Rembrandt Harmensz. van Rijn",
11  "longTitle": "Schutters van wijk II onder leiding van kapitein Frans Banninck  
Cocq, bekend als de 'Nachtwacht', Rembrandt Harmensz. van Rijn, 1642",
12  "showImage": true,
13  "permitDownload": true,
14  "webImage": {
15    "guid": "3ae88fe0-021c-41ae-a4ce-cc70b7bc6295",
16    "offsetPercentageX": 50,
17    "offsetPercentageY": 100,
18    "width": 2500,
19    "height": 2034,
20    "url":
21      "http://lh6.ggpht.com/ZYWwML8mVFonXzbmg2rQBulNuCSr3rAaf5ppNcUc2Id8qXqudDL1NSYxaqj  
EXyDLSbeNFz0HRu0H7rbIws0Js4d7s_M=s0"
22  }
23 }
```

line: 1 / 23 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: JSON scope: unknown

JSON: JavaScript Object Notation

JSON (Tweet)



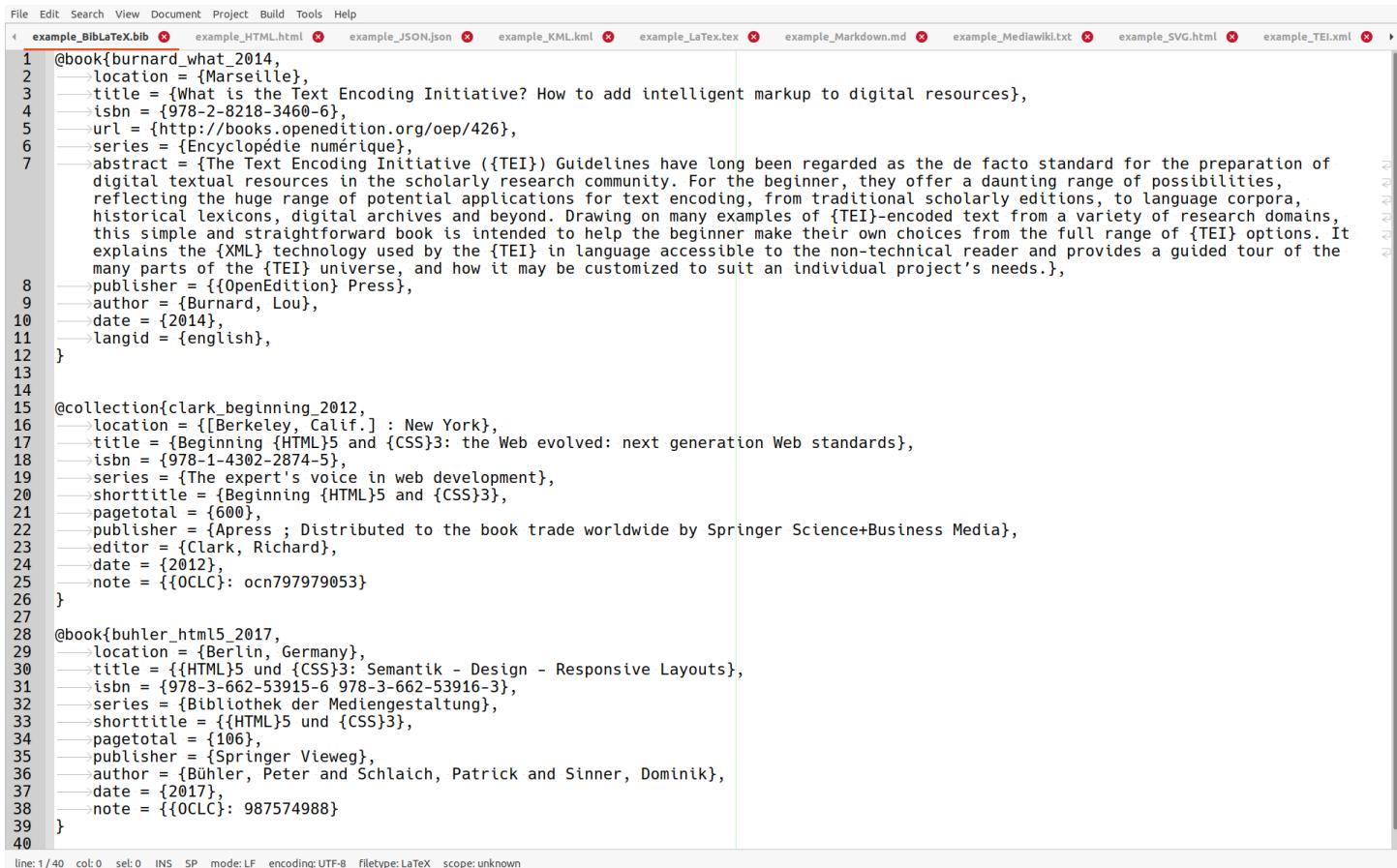
The screenshot shows a code editor interface with a tab bar at the top containing several file names: example_JSON.json, example_KML.kml, example_LaTeX.tex, example_Markdown.md, example_Mediawiki.txt, example_SVG.html, example_TEI.xml, example_TSV.tsv, and example_Tweet-in-JSON.json. The tab for 'example_Tweet-in-JSON.json' is highlighted with a red border. The main area of the editor displays the following JSON code:

```
1 {
2     "created_at": "Thu Apr 06 15:24:15 +0000 2017",
3     "id_str": "850006245121695744",
4     "text": "1\ufe0f Today we\u2019re sharing our vision for the future of the Twitter API\nplatform!\nhttps://t.co/XweGngmxlP",
5     "user": {
6         "id": 2244994945,
7         "name": "Twitter Dev",
8         "screen_name": "TwitterDev",
9         "location": "Internet",
10        "url": "https://dev.twitter.com/",
11        "description": "Your official source for Twitter Platform news, updates & events. Need\ntechnical help? Visit https://twittercommunity.com/\u2328\ufe0f #TapIntoTwitter"
12    },
13    "place": {},
14 },
15   "entities": {
16     "hashtags": [
17     ...
18     ],
19     "urls": [
20       {
21         "url": "https://t.co/XweGngmxlP",
22         "unwound": {
23             "url": "https://cards.twitter.com/cards/18ce53wgo4h/3xo1c",
24             "title": "Building the Future of the Twitter API Platform"
25         }
26       }
27     ],
28     "user_mentions": [
29     ]
30   }
```

At the bottom of the editor, there is a status bar with the following information: line: 1 / 35, col: 0, sel: 0, INS, SP, mode: LF, encoding: UTF-8, filetype: JSON, scope: unknown.

JSON (Tweet)

BibLaTeX



The screenshot shows a code editor window with multiple tabs at the top, all of which are currently visible. The active tab is 'example_BibLaTeX.bib'. The code in the editor is a BibLaTeX bibliography file containing entries for three books. The first entry is for a book by Burnard, Lou, published in 2014. The second entry is for a book by Clark, Richard, published in 2012. The third entry is for a book by Buhler, Peter and Schlaich, Patrick, published in 2017. Each entry includes fields for location, title, ISBN, series, abstract, publisher, author, date, and language.

```
File Edit Search View Document Project Build Tools Help
< example_BibLaTeX.bib ✘ example_HTML.html ✘ example_JSON.json ✘ example_KML.kml ✘ example_LaTeX.tex ✘ example_Markdown.md ✘ example_Mediawiki.txt ✘ example_SVG.html ✘ example_TEI.xml ✘ >
1 @book{burnard_what_2014,
2   location = {Marseille},
3   title = {What is the Text Encoding Initiative? How to add intelligent markup to digital resources},
4   isbn = {978-2-8218-3460-6},
5   url = {http://books.openedition.org/oep/426},
6   series = {Encyclopédie numérique},
7   abstract = {The Text Encoding Initiative (TEI) Guidelines have long been regarded as the de facto standard for the preparation of digital textual resources in the scholarly research community. For the beginner, they offer a daunting range of possibilities, reflecting the huge range of potential applications for text encoding, from traditional scholarly editions, to language corpora, historical lexicons, digital archives and beyond. Drawing on many examples of TEI-encoded text from a variety of research domains, this simple and straightforward book is intended to help the beginner make their own choices from the full range of TEI options. It explains the XML technology used by the TEI in language accessible to the non-technical reader and provides a guided tour of the many parts of the TEI universe, and how it may be customized to suit an individual project's needs.},
8   publisher = {{OpenEdition} Press},
9   author = {Burnard, Lou},
10  date = {2014},
11  langid = {english},
12 }
13
14
15 @collection{clark_beginning_2012,
16   location = {[Berkeley, Calif.] : New York},
17   title = {Beginning {HTML}5 and {CSS}3: the Web evolved: next generation Web standards},
18   isbn = {978-1-4302-2874-5},
19   series = {The expert's voice in web development},
20   shorttitle = {Beginning {HTML}5 and {CSS}3},
21   pagetotal = {600},
22   publisher = {Apress ; Distributed to the book trade worldwide by Springer Science+Business Media},
23   editor = {Clark, Richard},
24   date = {2012},
25   note = {{OCLC}: ocn797979053}
26 }
27
28 @book{buhler_html5_2017,
29   location = {Berlin, Germany},
30   title = {{HTML}5 und {CSS}3: Semantik - Design - Responsive Layouts},
31   isbn = {978-3-662-53915-6 978-3-662-53916-3},
32   series = {Bibliothek der Mediengestaltung},
33   shorttitle = {{HTML}5 und {CSS}3},
34   pagetotal = {106},
35   publisher = {Springer Vieweg},
36   author = {Buhler, Peter and Schlaich, Patrick and Sinner, Dominik},
37   date = {2017},
38   note = {{OCLC}: 987574988}
39 }
40
line: 1 / 40 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: LaTeX scope: unknown
```

BibLaTeX

Markup-basiert: XML

HTML (xHTML)

```
File Edit Search View Document Project Build Tools Help
< example_BibLaTeX.bib < example_HTML.html < example_JSON.json < example_KML.kml < example_LaTeX.tex < example_Markdown.md < example_Mediawiki.txt < example_SVG.html < example_TEI.xml >
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <title>Vorlesung</title>
5     <!-- link to main stylesheet -->
6     <link rel="stylesheet" type="text/css" href="/css/main.css">
7   </head>
8   <body>
9     <nav>
10       <ul>
11         <li><a href="/">Home</a></li>
12         <li><a href="/vorlesung">Vorlesung</a></li>
13         <li><a href="/projekte">Projekte</a></li>
14         <li><a href="/talks">Präsentationen</a></li>
15         <li><a href="/links">Links</a></li>
16       </ul>
17     </nav>
18     <div class="container">
19       
20       <div class="blurb">
21         <h1>Vorlesung: Einführung in die Digital Humanities</h1>
22         <p><a href="https://creativecommons.org/licenses/by/4.0/"><img src=
23           "https://img.shields.io/badge/License-CC%20BY%204.0-lightgrey.svg"></img></a></p>
24         <ul>
25           <li>E01: Was sind die Digital Humanities: <a href=
26             "https://dh-trier.github.io/dh-vorlesung/reveal/DH-E01_Was-ist-Digital-Humanities.html">online</a> | <a href=
27             "https://github.com/dh-trier/dh-vorlesung/raw/gh-pages/pdf/DH-E01_Was-ist-Digital-Humanities.pdf">pdf</a></li>
28           <li>E02: Digitalisierung: Text und Bild: <a href=
29             "https://dh-trier.github.io/dh-vorlesung/reveal/DH-E02_Digitalisierung.html">online</a> | <a href=
30             "https://github.com/dh-trier/dh-vorlesung/raw/gh-pages/pdf/DH-E02_Digitalisierung.pdf">pdf</a></li>
31           <li>E03: Grundbegriffe des Programmierens: <a href=
32             "https://dh-trier.github.io/dh-vorlesung/reveal/DH-E03_Programmieren.html">online</a> | <a href=
33             "https://github.com/dh-trier/dh-vorlesung/raw/gh-pages/pdf/DH-E03_Programmieren.pdf">pdf</a></li>
34         </ul>
35       </div>
36     </div>
37   </body>
38 </html>
```

line: 1 / 33 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: HTML scope: unknown

HTML: HyperText Markup Language

XML (WLV)

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="wlv-schema.rnc" type="application/relax-ng-compact-syntax"?>
<wlv id="smw" url="https://github.com/dh-trier/wlv/smw">
  <metadata>
    <title>Sammlung M. W. (SMW)</title>
    <editor ref="wikidata:Q57917455">Mike Mustermann</editor>
    <institution url="https://kompetenzzentrum.uni-trier.de">Trier Center for Digital Humanities</institution>
    <address ref="wikidata:Q3138">Trier</address>
    <date>2020</date>
    <licence licenceScope="metadata" licenceAbbr="CC0" url="https://creativecommons.org/publicdomain/zero/1.0/deed.en">Public Domain Dedication</licence>
    <licence licenceScope="textual">public domain</licence>
    <licence licenceScope="visual">to be determined.</licence>
    <curators>
      <curator id="mm" ref="wikidata:Q57917455">Maria Musterfrau</curator>
    </curators>
  </metadata>
  <labels>
    <label curatorID="#mm" labelType="Gutsetikett" curationDate="2020-08-29" id="smw-0001">
      <labelPart partType="front" partNum="1">
        <physical sizeH="110" sizeV="69" shape="other" material="paper" printingTechnique="litho+type"/>
        <visual>
          <background backgroundStyle="color">white</background>
          <frame frameStyle="pattern" frameColor="light blue" frameType="outer"/>
          <frame frameStyle="floral" frameColor="grey" frameType="outer"/>
          <figure figureType="realistic" figureNum="1" figurePosition="upper-center">Grapes</figure>
        </visual>
        <textual>
          <wineOrigin wineOriginType="vineyard (Lage)" wineOriginNorm="Dhron" url="https://www.riesling.de/de/lage/dhroner-hofberger/53_184.html">Thronhofberger</wineOrigin>
          <agent agentRole="distributor/importer/exporter">E. F. Cronenbold</agent>
          <location locationRole="distributor/importer/exporter" locationType="locality" ref="wikidata:Q3138">Trier</location>
        </textual>
      </labelPart>
      <provenance>
        <dating notBefore="1850" notAfter="1900" certainty="low" curatorID="#cs"/>
      </provenance>
    </label>
  </labels>
</wlv>
```

line: 1 / 39 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: XML scope: unknown

XML: eXtensible Markup Language

KML



The screenshot shows a code editor window with a menu bar at the top: File, Edit, Search, View, Document, Project, Build, Tools, Help. Below the menu is a tab bar with several tabs: example_BibLaTeX.bib, example_HTML.html, example_JSON.json, example_KML.kml, example_LaTeX.tex, example_Markdown.md, example_Mediawiki.txt, example_SVG.html, and example_TEI.xml. The example_KML.kml tab is currently selected. The main editor area contains the following KML XML code:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <kml xmlns="http://www.opengis.net/kml/2.2">
3   <Placemark>
4     <name>Trier Center for Digital Humanities</name>
5     <description>Forschungszentrum für digitale Geisteswissenschaften an der  
Universität Trier.</description>
6     <Point>
7       <coordinates>49.7744575,6.684948</coordinates>
8     </Point>
9   </Placemark>
10 </kml>
11
```

At the bottom of the editor, there is a status bar with the following information: line: 1 / 11, col: 0, sel: 0, INS, SP, mode: LF, encoding: UTF-8, filetype: XML, scope: unknown.

KML: Keyhole Markup Langauge
siehe auch: GML (Geography Markup Language)

TEI

```
<?xml version="1.0" encoding="utf-8"?>
<?xml-model href="../../Schemas/eltec-x.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="../../Schemas/eltec-x.rng" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xml:id="Bernardin_Paul" xml:lang="fr" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titlestmt>
        <title ref="bgrf:1">Paul et Virginie:MiMoText edition</title>
        <author ref="viaf:9966915;wikidata:Q3261946">Jacques Henri Bernardin de Saint-Pierre(1737-1814)</author>
        <respstmt>
          <resp>data capture</resp>
          <name>Ebooks libres et gratuits contributors</name>
        </respstmt>
        <respstmt>
          <resp>encoding</resp>
          <name>Julia Duda</name>
        </respstmt>
        <respstmt>
          <resp>editor</resp>
          <name>Julia Röttgermann</name>
        </respstmt>
      </titlestmt>
      <extent>
        <measure unit="words">60490</measure>
      </extent>
      <publicationstmt>
        <publisher ref="https://mimotext.uni-trier.de">Mining and Modeling Text</publisher>
        <distributor ref="https://github.com/mimotext/roman-dixhuit">Github</distributor>
        <date>2020</date>
        <availability>
          <licence target="https://creativecommons.org/publicdomain/zero/1.0/deed.en"/>
        </availability>
      </publicationstmt>
      <sourceDesc>
        <bibl type="digitalSource">
          <ref target="https://www.ebooksgratuits.com/details.php?book=1682"/>
        </bibl>
        <bibl type="printSource">
          <date>1787</date>
        </bibl>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <body>
    <!-- Content of the book -->
  </body>
</TEI>
```

TEI: Text Encoding Initiative

SVG



The screenshot shows a code editor interface with a menu bar at the top. The menu items include File, Edit, Search, View, Document, Project, Build, Tools, and Help. Below the menu, there is a tab bar with several tabs: example_BibLaTeX.bib, example_HTML.html, example_JSON.json, example_KML.kml, example_LaTeX.tex, example_Markdown.md, example_Mediawiki.txt, example_SVG.html (which is currently selected), and example_TEI.xml. The main editor area displays the following HTML code:

```
1 <!DOCTYPE html>
2 <html>
3 <body>
4
5 <svg height="210" width="500">
6   <polygon points="100,10 40,198 190,78 10,78 160,198"
7     style="fill:lime;stroke:purple;stroke-width:5;fill-rule:nonzero;" />
8 </svg>
9 </body>
10 </html>
11
```

At the bottom of the editor, there is a status bar with the following information: line: 1 / 11, col: 0, sel: 0, INS, SP, mode: CRLF, encoding: UTF-8, filetype: HTML, scope: unknown.

SVG: Scalable Vector Graphics (live)

Markup-basiert: nicht XML

LaTeX

```
File Edit Search View Document Project Build Tools Help
< example_BibLaTeX.bib > example_HTML.html > example_JSON.json > example_KML.kml > example_LaTeX.tex < example_Markdown.md > example_Medawiki.txt > example_SVG.html > example_TEI.xml >
1 \title{Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen}
2
3 \author{
4 Christof Schöch <schoech@uni-trier.de>\\
5 Frédéric Döhl <f.doehl@dnb.de>\\
6 Achim Rettlinger <rettlinger@uni-trier.de>\\
7 Evelyn Gius <evelyn.gius@tu-darmstadt.de>\\
8 Peer Trilcke <trilcke@uni-potsdam.de>\\
9 Peter Leinen <P.Leinen@dnb.de>\\
10 Fotis Jannidis <fotis.jannidis@uni-wuerzburg.de>\\
11 Maria Hinzmann <hinzmannm@uni-trier.de>\\
12 Jörg Röpke <roepke@uni-trier.de>\\
13 }
14
15 \begin{document}
16 \maketitle
17 \section{Zusammenfassung}
18 Das Text und Data Mining (TDM) mit urheberrechtlich geschützten Texten unterliegt trotz der TDM-Schranke (§60d UrhG) weiterhin Einschränkungen, die u.a. die Speicherung, Veröffentlichung und Nachnutzung der entstehenden Korpora betreffen und das volle Potenzial des TDM in den Digital Humanities ungenutzt lassen. Als Lösung werden abgeleitete Textformate vorgeschlagen: Hier werden urheberrechtlich geschützte Textbestände so transformiert, dass alle wesentlichen urheberrechtlich relevanten Merkmale entfernt werden, verschiedene einschlägige Methoden des TDM aber möglich bleiben. Mehrere abgeleitete Textformate werden aus Sicht der Computational Literary Studies, der Informatik, der Gedächtnisinstitutionen und der Rechtswissenschaften beleuchtet.
19
20 \section{1. Einleitung}
21
22 Es ist ein offenes Geheimnis in den Digital Humanities (DH), dass es für die Computational Literary Studies (CLS) bezüglich der verfügbaren Textbestände ein \textit{window of opportunity} gibt, das sich um 1800 öffnet und um 1920 wieder schließt. Es öffnet sich erst um 1800, weil für Materialien vor dieser Zeit die technischen Herausforderungen im Bereich Optical Character Recognition (OCR) und Normalisierung von orthographischer Varianz immer noch so groß sind, dass deutlich weniger umfangreiche bzw. qualitativ weniger hochwertige Textsammlungen zur Verfügung stehen als für die Zeit nach 1800. Und es schließt sich um 1920, weil für Materialien, die später erschienen sind, in sehr vielen Fällen (abhängig vom Todesdatum der Autor/innen) das Urheberrecht nach wie vor greift und sowohl das Erstellen als auch das Teilen von Textsammlungen mit Dritten deutlich erschwert. Dieser Umstand hat bedauerlicherweise zur Folge, dass die Setzung von Forschungsschwerpunkten häufig nicht primär von den Erkenntnisinteressen und Zielen der Forschung selbst
```

line: 1 / 34 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: LaTeX scope: unknown

LaTeX

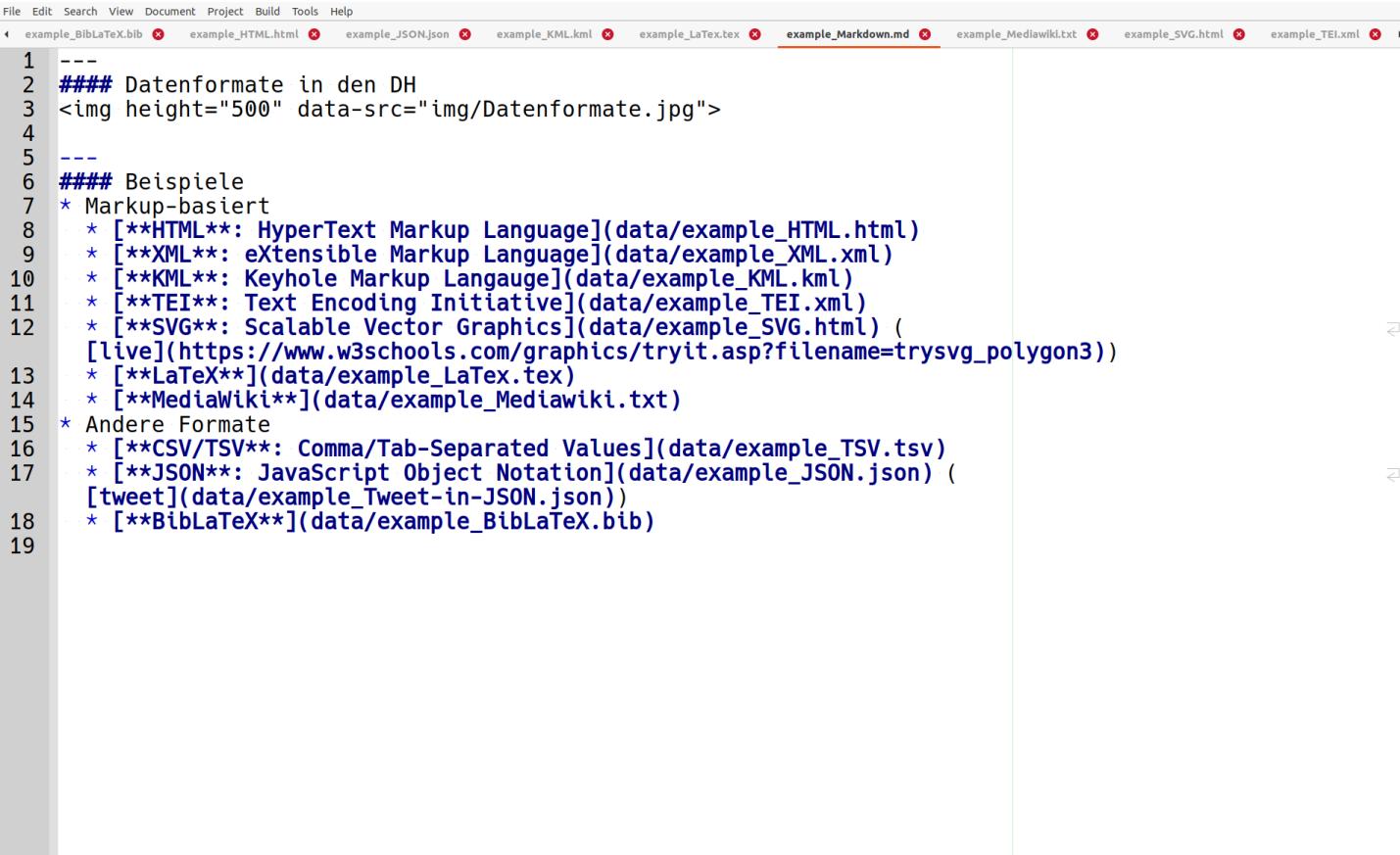
Mediawiki

```
File Edit Search View Document Project Build Tools Help
< example_BibLaTeX.bib example_HTML.html example_JSON.json example_KML.kml example_LaTeX.tex example_Markdown.md example_Mediawiki.txt example_SVG.html example_TEI.xml >
1 Die ''''Universität Trier''' ist eine mittelgroße staatliche [[Universität]] in der
[[Rheinland-Pfalz|rheinland-pfälzischen]] [[Stadt]] [[Trier]] an der [[Mosel]]. Gegründet im Jahr
1473 und 1798 unter französischer Besatzung nach 325-jährigem Bestehen geschlossen, wurde sie 1970
neu eröffnet. Etwa 12.500 Studierende und insgesamt rund 2.000 Beschäftigte lernen, arbeiten und
forschen auf dem Campus in Höhenlage über der Stadt. Im Jahr 2013 wurde das Universitätsgelände
vom [[Bund Deutscher Landschaftsarchitekten]] als einer von 100 besonderen Orten in Deutschland
ausgezeichnet.
2
3 == Geschichte ==
4
5 === Erstgründung ===
6
7 Der Entschluss zur Gründung der Universität Trier dürfte im Jahr 1450 gefasst worden sein
(sogenanntes „Heiliges Jahr“), als der [[Kurtrier|Trierer Kurfürst-Erzbischof]] [[Jakob I. von
Sierck]] nach Rom pilgerte und mit [[Nikolaus von Kues]] zusammentraf, der dort als
[[Kurienkardinal]] tätig war und dem Gedanken einer Universitätsgründung in seiner Heimatregion
vermutlich aufgeschlossen gegenüberstand. Papst [[Nikolaus V. (Papst)|Nikolaus V.]] erlaubte
daraufhin am 2. Februar 1455 dem Trierer Erzbischof die Gründung einer Universität. Jakob von
Sierck konnte jedoch aus finanziellen Gründen das Vorhaben nicht ausführen und starb auch kurze
Zeit nach Erlass der entsprechenden päpstlichen Bullen. Sein Nachfolger [[Johann II. von Baden]]
war in diverse [[Fehde]]n verwickelt und hatte dadurch weder Geld noch Zeit, sich um eine
Universitätsgründung in Trier zu kümmern.
8
9 == Lage ==
10
11 Die Universität liegt auf einem Plateau rund 150 Höhenmeter oberhalb der Innenstadt zwischen den
Trierer Stadtteilen [[Trier-Kürenz|Kürenz]], [[Trier-Tarforst|Tarforst]] und
[[Trier-Olewig|Olewig]]. Zum Universitätsgelände des Campus I gehören ausgedehnte Park- und
Grünanlagen mit Spazierwegen, Teichen, landschaftlichen Skulpturen und Plastiken. Die Universität
befindet sich in direkter Nachbarschaft zum Gelände der Trierer [[Landesgartenschau]] 2004.
12
13 Studiengangsbeschreibung
```

line: 1 / 49 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: None scope: unknown

MediaWiki

Markdown



The screenshot shows a code editor interface with a menu bar at the top. Below the menu is a tab bar with several tabs, one of which is highlighted in red and labeled "example_Markdown.md". The main area of the editor displays the following text:

```
1 ---  
2 ##### Datenformate in den DH  
3   
4 ---  
5 ##### Beispiele  
6 * Markup-basiert  
7   * [**HTML**: HyperText Markup Language](data/example_HTML.html)  
8   * [**XML**: eXtensible Markup Language](data/example_XML.xml)  
9   * [**KML**: Keyhole Markup Language](data/example_KML.kml)  
10  * [**TEI**: Text Encoding Initiative](data/example_TEI.xml)  
11  * [**SVG**: Scalable Vector Graphics](data/example_SVG.html) ( [live](https://www.w3schools.com/graphics/tryit.asp?filename=trysvg_polygon3) )  
12  * [**LaTeX**](data/example_LaTex.tex)  
13  * [**MediaWiki**](data/example_Mediawiki.txt)  
14 * Andere Formate  
15   * [**CSV/TSV**: Comma/Tab-Separated Values](data/example_TSV.tsv)  
16   * [**JSON**: JavaScript Object Notation](data/example_JSON.json) ( [tweet](data/example_Tweet-in-JSON.json) )  
17   * [**BibLaTeX**](data/example_BibLaTeX.bib)  
18  
19
```

At the bottom of the editor window, there is a status bar with the following information: line: 1 / 19 col: 0 sel: 0 INS SP mode: LF encoding: UTF-8 filetype: Markdown scope: unknown.

Markdown

Fazit

Fazit Auszeichnungssprachen

Fazit Auszeichnungssprachen

- Vorteile

- Von Menschen und Maschinen lesbar
- Metadaten, Markup und Text/Daten bleiben zusammen
- Inhalt und Struktur gemeinsam, Darstellung davon getrennt (CSS)
- Daten sind letztlich einfach Textdateien (nachhaltig)

Fazit Auszeichnungssprachen

- Vorteile

- Von Menschen und Maschinen lesbar
- Metadaten, Markup und Text/Daten bleiben zusammen
- Inhalt und Struktur gemeinsam, Darstellung davon getrennt (CSS)
- Daten sind letztlich einfach Textdateien (nachhaltig)

- Nachteile

- Etwas "verbose"
- komplexer zu erstellen als "plain text"
- weniger performant als Datenbanken

Fazit XML

Fazit XML

- Vorteile

- XML ist eine Meta-Auszeichnungssprache (siehe: SVG, KML, TEI, etc.)
- Man kann damit beliebige Auszeichnungssprachen definieren (siehe: WLV)
- Validierung nach Schema (Datenmodell) möglich (sehr wichtig!)
- für Textauszeichnung und Datenhaltung geeignet
- extrem weit verbreitet

Fazit XML

- **Vorteile**
 - XML ist eine Meta-Auszeichnungssprache (siehe: SVG, KML, TEI, etc.)
 - Man kann damit beliebige Auszeichnungssprachen definieren (siehe: WLV)
 - Validierung nach Schema (Datenmodell) möglich (sehr wichtig!)
 - für Textauszeichnung und Datenhaltung geeignet
 - extrem weit verbreitet
- **Nachteile**
 - (wie für Auszeichnungssprachen allgemein)

Lektüreempfehlung

- Kapitel "Datenformate" in: *IT-Empfehlungen. Für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften*, IANUS, Version 1.0.1.0, 2017. <https://www.ianus-fdz.de/it-empfehlungen/>.
- Johanna Puhl (Hauptautorin): "Empfehlungen für Forschungsdaten, Tools und Metadaten in der Dariah-DE Infrastruktur", *Dariah-DE Wiki*, 2017, <https://wiki.de.dariah.eu/pages/viewpage.action?pageId=38080370>.



Danke!

Lizenz: Creative Commons Attribution (CC BY), 2020.
