

Semantic Roundtripping for internal Processes and cross-company Data Integration

Cornelsen

Cornelsen in numbers

The publisher Cornelsen is one of the leading providers of educational media in German-speaking countries.

More than 1300 employees work at our head office in Berlin and seven other locations.

1946



Publishing house founded
We have been active in education for more than 70 years

1300



Employees
at Cornelsen Verlag and at the Franz Cornelsen education holding company

22



Nationalities
getting on fantastically well at our offices

5.000



Authors and editors
work with us on inspiring education solutions

53



Professional titles
Under one roof – in addition to publishers, there are scrum masters, business analysts and developers

80



per cent
of all German teachers use Cornelsen products in their work

1:1



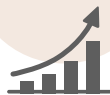
Half
of all managers are female

19.000



Partners
are part of our global network from booksellers to the Goethe Institute

23.000



Titles
And the publishing programme is growing from year to year

4,8 Mio.



Textbooks
sold including access codes for e-books

10.000



teaching materials
Available for download

7,8 Mio.



Every pupil
in Germany owns at least one Cornelsen workbook

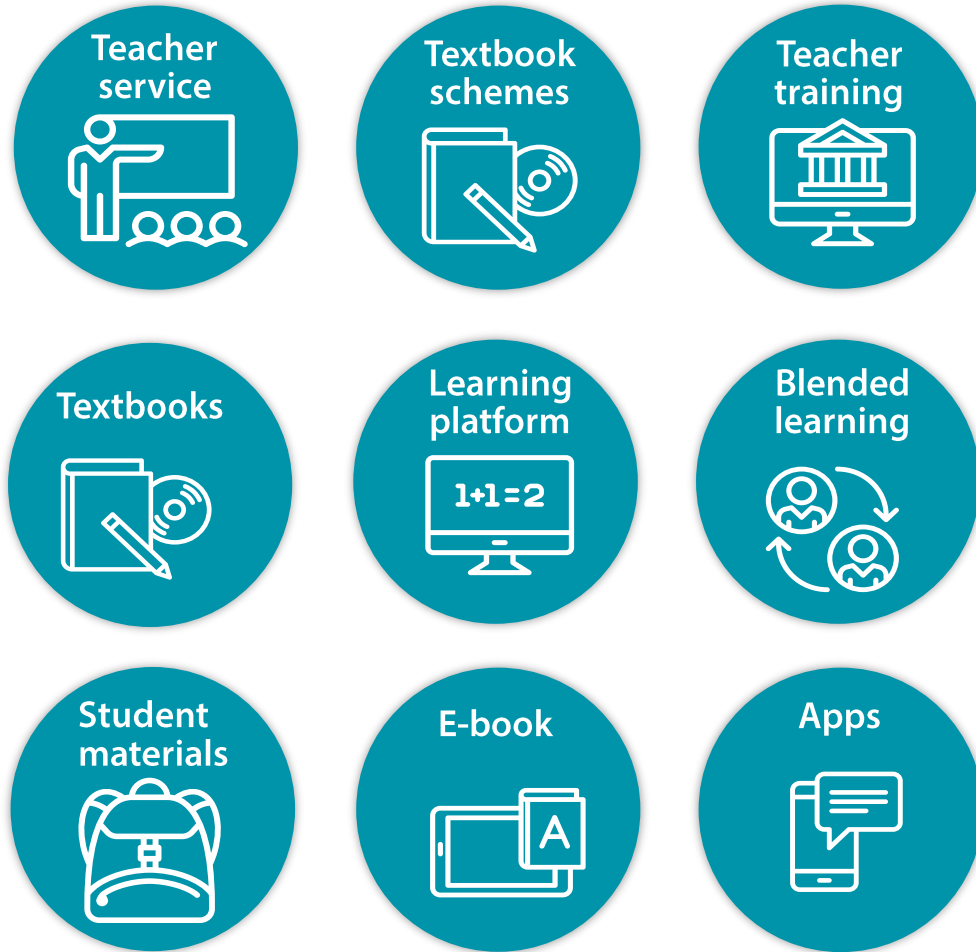
Teachers and pupils trust in our brands

Teachers and pupils trust us because Cornelsen's wide-ranging learning methods and learning media are developed in close cooperation with both parties.

The range is shaped by everyday school life, relieving the pressure on teachers and encouraging effective and lively lessons.



Welcome to the world of learning



The publishing programme comprises more than 23,000 titles for all subjects, school types and German states – from early years education to teaching and learning systems for secondary schools and vocational schools, and educational media for adult education and specialist pedagogical literature.

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

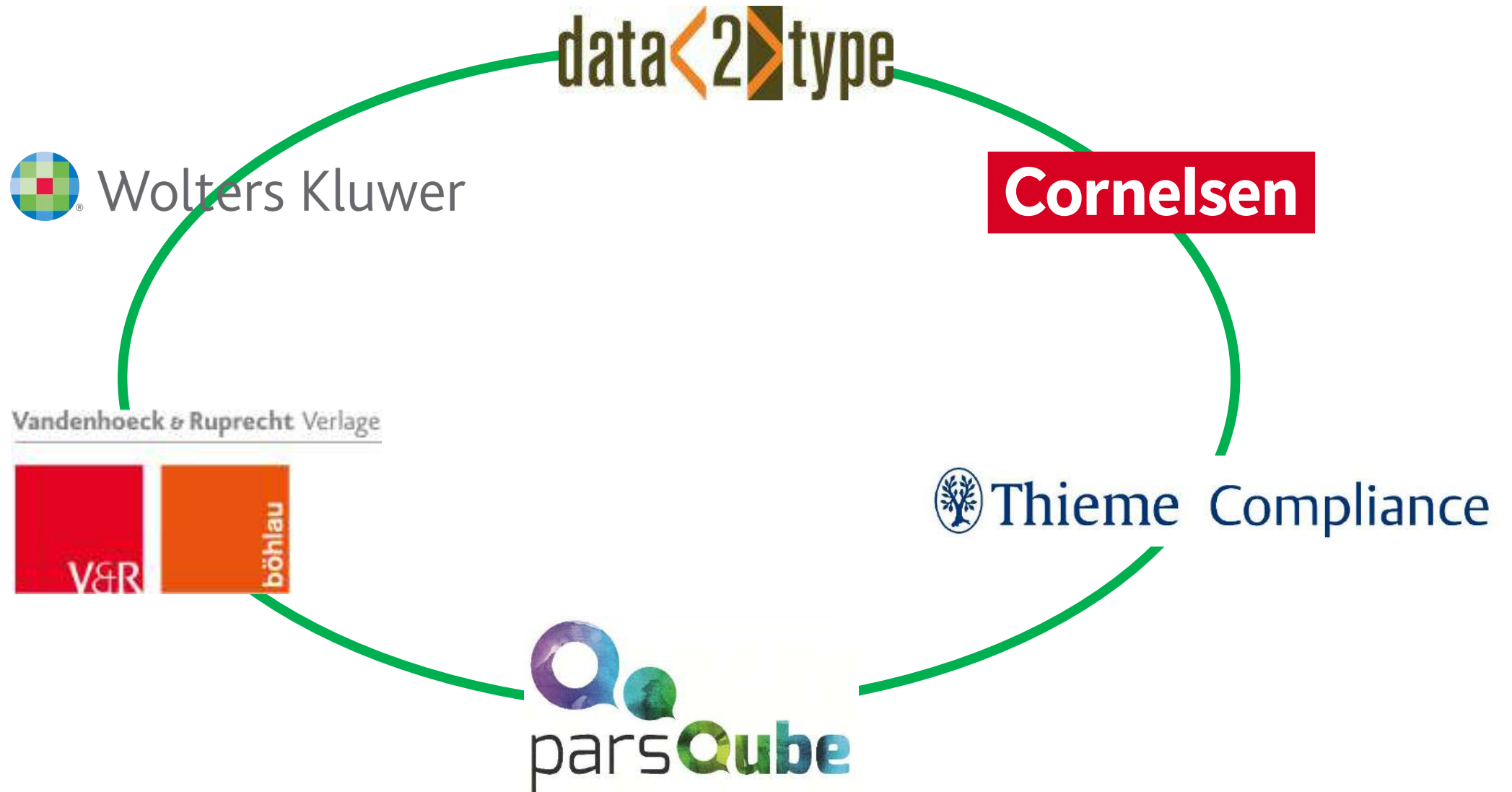
Goals of our project

- A great variety of existing industries are transformed by semantic technologies
- In these industries, many processes are based on XML
- XML is not just XML!
 - Each industry uses (and sometimes develops) dozens or even hundreds of XML formats
- XML usage and interplay between formats is always embedded in (company internal) processes
- Basic hypothesis: Understanding the role and interplay of XML formats in selected processes will lead to new business opportunities for semantic technologies
 - Bring added value to existing processes
 - Avoid the need to change processes
 - Allow to do semantic processing without breaking formats & processes
- Goal of this project: validate this hypothesis in dedicated use cases

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

Project Partners



Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

Wolters Kluwer Use Case: Automatic content enrichment

Task

- Detect simple entities like names and simple content structures like legal paragraphs in unstructured text
- Detect complex content structures like legal facts in court decisions
- Feed the detected data back into existing XML files or as RDF in the CMS triple store

Goal

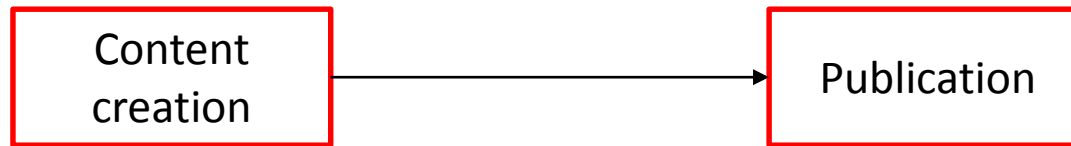
- Improve the existing and very efficient XML workflow with additional enrichment capabilities using ML and NLP, so that the process is further simplified, faster and therefore also more cost effective
- This will help to broaden the applicability of semantic web technologies within Wolters Kluwer

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

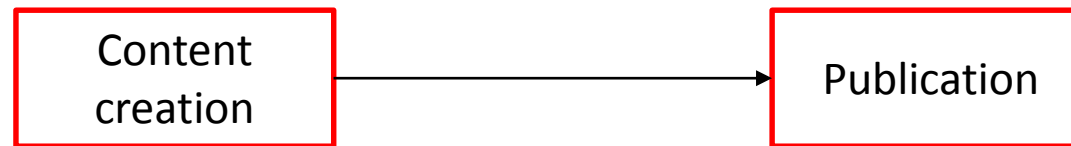
Cornelsen use case: automatic tagging

Birds eye view on educational publishing



Cornelsen use case: automatic tagging

Reality – a short list of *example* processes involved

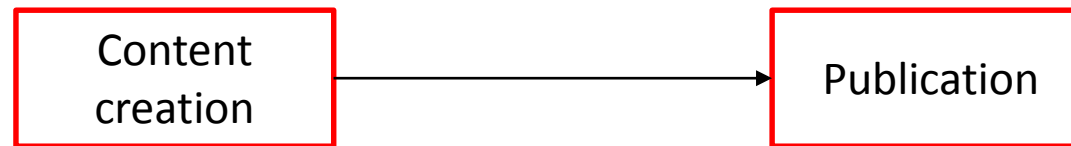


- Analysis of existing content
- Content model creation
- Content model re-use
- Content transformation
- Transformation output quality assurance (manual + automatic)
- Re-transformation
- Editing environment parametrization
- Content Editing
- Output preview
- Quality assurance
- Digital rights management verification
- Pre-publication
- Product testing
- Actual publication

Cornelsen use case: automatic tagging

Automatic tagging: enrich content with links to internal or public data sources

- Potentially relevant for **several processes**



- Analysis of existing content
- **Content model creation**
- Content model re-use
- **Content transformation**
- **Transformation output quality assurance** (manual + automatic)
- **Re-transformation**
- **Editing environment parametrization**
- **Content Editing**
- **Output preview**
- **Quality assurance**
- Digital rights management verification
- Pre-publication
- **Product testing**
- Actual publication

Cornelsen use case: automatic tagging

- Semantic technologies without XML integration: several processes need to be changed
 - Content model creation, content transformation, transformation output quality assurance, re-transformation, editing environment parametrization, content editing, output preview, quality assurance, product testing
- Cornelsen uses the format
 - DITA
 - The Learning & Training Content Specialization
 - With Cornelsen specific adaptations
- Main goal of this use case: integrate automatic tagging into DITA L&T CV processes, while avoiding large adaptations to these processes

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

Demo: combining XML and RDF processing

Geschichte.docx [Kompatibilitätsmodus] - Word

Ansicht ACROBAT Was möchten Sie tun? Anmelden Freigeben

Lesemodus Seitenlayout Weblayout

Gliederung Entwurf

Lineal Gitternetzlinien Navigationsbereich

Zoom 100%

Eine Seite Mehrere Seiten Seitenbreite

Neues Alle Teilen Fenster anordnen

Nebeneinander anzeigen Synchrones Scrollen Fensterposition zurücksetzen

Fenster wechseln Makros


40 2 Ägypten - eine frühe Hochkultur.

Wer regierte Ägypten?

Das Wort Pharaon bedeutet „großer Haus“ und bezeichnete den Palast des Königs. Die ägyptischen Herrscher nannten sich zur Zeit der Pyramidenbau „Herrscher von Ober- und Unterägypten“, oder „Herrscher der beiden Länder“, in späteren Zeiten Pharaonen.

- Welche Macht löste ein Pharaon aus und welche Aufgaben musste er erfüllen?

Wie regierte der Pharaon?
Der Pharaon stand an der Spitze des Staates. Seit etwa 2500 v. Chr. wurde ein ägyptischer König als Sohn des Sonnengottes Ptahhotep verehrt. Er galt dabei als heiliges und unsterbliches göttliches Wesen. Der Pharaon regierte über das gesamte Land. Ihm gehörten alle Lande an. Als Herrscher musste er die Gesetze und setzte die Beamten ein. Er war Richter über Leben und Tod, oberster Kriegsherr und oberster Priester. In seinem Auftrag wurden Pyramiden, Tempel und Städte gebaut. Seine Hauptaufgabe war die Sicherung des Friedens nach innen und außen. Der Pharaon lebte mit seiner Hausfrau und zehntausend Nebenfrauen, abgesehen von seinem Volk in einer großen Palastanlage. Im Palast wohnten auch Priester, Beamte, Musikanten, Tänzerinnen, Köche und viele Dienerinnen und Diener. Starb ein Pharaon, herrschte 70 Tage Toten im Totenland und bildeten ägyptische Herrscher danach die Zeitrechnung wieder neu.




Der Pharaon Ramses II. als Beschützer seines Volkes und seines Landes gegen die Bedrohung durch Kriegerstämme in Abu Simbel, um 1250 v. Chr.

Ein Relief aus dem Tempel von Karnak, das den Pharaon Ramses II. zeigt, wie er die Götter Osiris und Isis verehrt. Um 1250 v. Chr.

Grundlegender Begriff: Monarchie
Ein Staat, in dem die Regierung in einer Monarchie (ein Herrscher) liegt. Der Herrscher (Pharaon, König, Kaiser) ist die höchste Macht im Staat. Er erlässt Gesetze und ist für das Wohlergehen seines Landes und den Schutz der Untertanen verantwortlich. Die Macht des Monarchen kann aus einem geschriebenen Urkunde entstehen, wird aber auch durch die Abstammung oder einen göttlichen Auftrag begründet.

41

So begrüßten die Ägypter den Pharaon Ramses II. (Regierungszeit 1279-1213 v. Chr.)
Wir kommen zu dir, Herr des Himmels, Herr der Erde, du lebende Sonne des ganzen Landes, Herr der Lebenden, du Sonnengott der Menschheit, du Staat des Himmels, du Balken der Erde. Hier ist die Sprache der Wahrheit, die du sprachst, wenn alle schickst, dessen Kraft Ägypten erstirbt, der über die Fremdlinger siegt und triumphierend heimkehrt, dessen Stärke Ägypten schützt. Geliebter der Wahrheit, der in seinen Gesetzen die Welt liebt, dessen Schrecken die Fremdlinger weichen lässt, du unser König, unser Herr.



Ein Relief aus dem Tempel von Karnak, das den Pharaon Ramses II. zeigt, wie er die Götter Osiris und Isis verehrt. Um 1250 v. Chr.

Grundlegender Begriff: Pharaon
(Mehrzahl: Pharaonen) oberster Herrscher im alten Ägypten, der als Gottkönig (Gott und König) verehrt wurde.

Der Ägyptologe Christian Jacq über den Tageslauf eines Pharaons
Der Pharaon konnte nicht zu und lassen, was er wollte. Er musste sich einem strengen Zeitplan unterwerfen (mit dem Ritual, Essen) des Sonnengottes besetzen. Der Pharaon am Morgen in den ersten Teil des Tages und gab die göttliche Statue aus einem abstrakten Schrein, der mit symbolischen Bildern von verschiedenen Göttern besetzt war. Er musste sich waschen und anziehen. Er durfte nicht essen, bis er sich waschen und anziehen würde. Der Pharaon trug die heilige Statue eines Gottes in seinem Boot (siehe die Abbildung). Er musste sich waschen und anziehen. Er durfte nicht essen, bis er sich waschen und anziehen würde. Der Pharaon trug die heilige Statue eines Gottes in seinem Boot (siehe die Abbildung). Er musste sich waschen und anziehen. Er durfte nicht essen, bis er sich waschen und anziehen würde. Der Pharaon trug die heilige Statue eines Gottes in seinem Boot (siehe die Abbildung).

1. Einleitung mittels von M1 und der Begriffsanwendung „Pharaon“ die verschiedenen Titel für die ägyptischen Könige.

2. Welche Aufgaben hatte ein Pharaon?
a) Beschützer seines Landes (Kriegsherr), welche Aufgaben und welche Mächten Pharaon hatte.
b) Arbeitsauftrag, welche Dienerinnen, die dem Pharaon Ramses II. von den Ägyptern zugeordnet wurden.

3. Partnerarbeit Vorgehensweise: Ein Partner liest den Text vor, der andere Partner liest den Text nach und beantwortet die Fragen.

4. a) Partnerarbeit Vorgehensweise: Ein Partner liest den Text vor, der andere Partner liest den Text nach und beantwortet die Fragen.

Seite 10 von 11 3336 Wörter 69 %

Demo: combining XML and RDF processing

The screenshot shows a Microsoft Word document titled "Geschichte_test.docx - Word". The ribbon includes tabs for "Datei", "Start", "Einfügen", "Entwurf", "Layout", "Verweise", "Sendungen", "Überprüfen", "Ansicht", "ACROBAT", and "Freigeben". The "Ansicht" tab is active, showing options for "Lesemodus", "Seitenlayout", and "Weblayout". The document content is organized into a grid of blocks:

- Section 40: Ägypten - eine frühe Hochkultur**
 - Wer regierte Ägypten?**

Das Wort *Pharao* bedeutet „großer König“, weil ausschließlich dem *Pharao* die göttlichen Zerschneidungsmächte zukamen. Er war der Herrscher über das Land, die Götter, die Menschen und die Tiere. Er war der Herrscher über das Land, die Götter, die Menschen und die Tiere. Er war der Herrscher über das Land, die Götter, die Menschen und die Tiere.
 - Person** (Ramesses II.)
 - um 1303 v. Chr.
 - † 27. Juni 1213 v. Chr.
 - https://de.wikipedia.org/wiki/Rammes_II
 - Bildatbank** (Statue of a pharaoh)
 - https://de.wikipedia.org/wiki/Tempel_von_Abu_Si_mel
- Section 41: Religion**
 - Ort** (Tempel)
 - <https://de.wikipedia.org/wiki/Tempel>
 - Bildatbank** (Temple interior)
 - <https://de.wikipedia.org/wiki/Tempel>
 - Bildatbank** (Temple exterior)
 - <https://de.wikipedia.org/wiki/Tempel>
 - Aufgaben**






Beachte, ob diese Herrschaftsform nur Vorteile für die ägyptische Bevölkerung mit sich brachte.



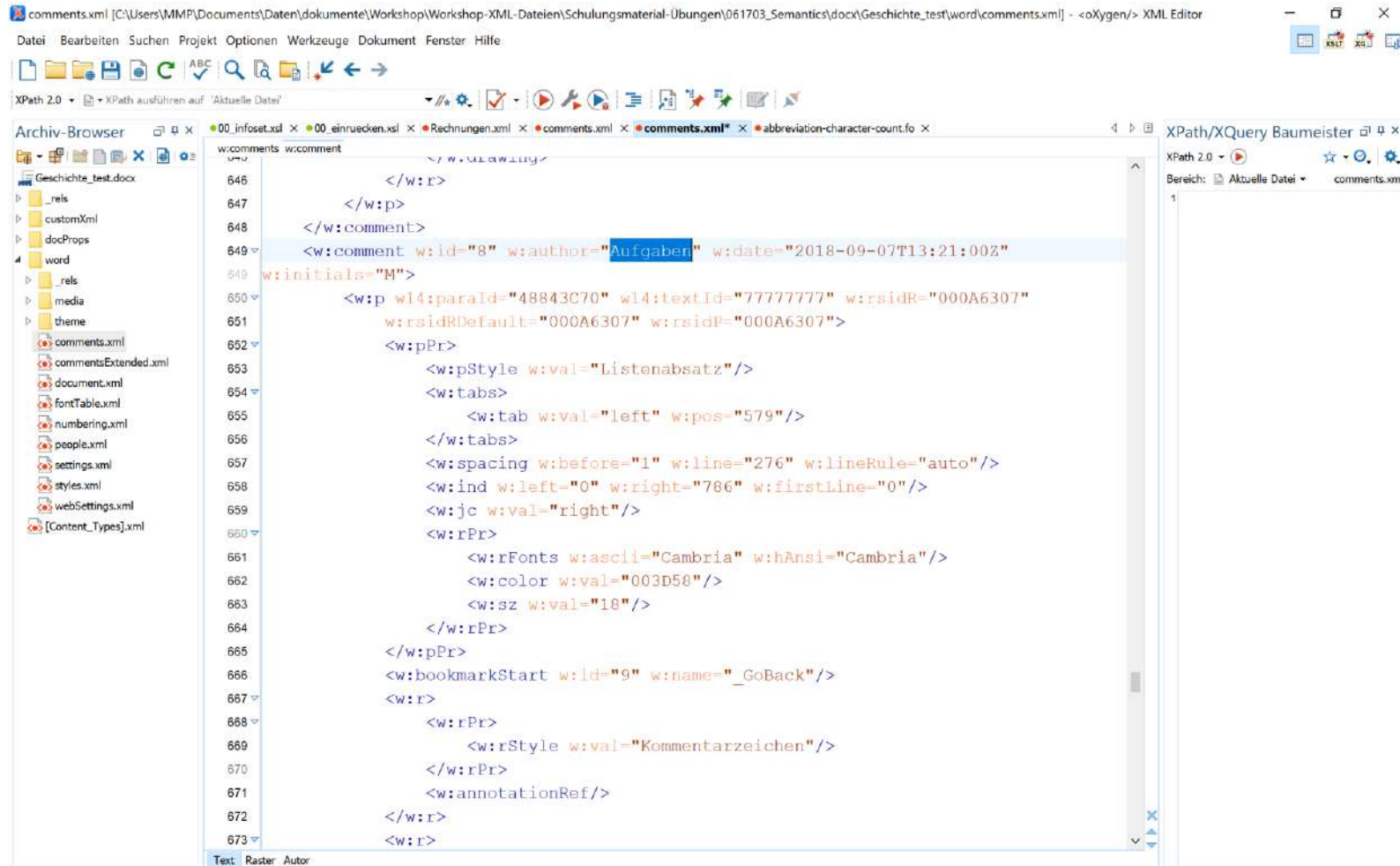
Demo: combining XML and RDF processing

Technologies

We work exclusively with XML technologies and use standards such as XSLT, Schematron, XPath, OOXML (Office Open XML).

 _rels	10.09.2018 16:56	Dateiordner	
 customXml	10.09.2018 16:56	Dateiordner	
 docProps	10.09.2018 16:56	Dateiordner	
<input checked="" type="checkbox"/>  word	10.09.2018 16:56	Dateiordner	
 [Content_Types].xml		XML Document	3 KB

Demo: combining XML and RDF processing



The screenshot shows the XML Editor interface with the following components:

- File Explorer (Archiv-Browser):** Shows a tree view of files including 'Geschichte_test.docx', '_rels', 'customXml', 'docProps', 'word', and various XML files like 'comments.xml', 'document.xml', 'fontTable.xml', 'numbering.xml', 'people.xml', 'settings.xml', 'styles.xml', 'webSettings.xml', and '[Content_Types].xml'.
- Main Editor:** Displays XML code for a comment. The code includes a root element <w:r>, followed by <w:p>, </w:p>, </w:comment>, and a <w:comment> element with attributes: w:id="8", w:author="Aufgaben", w:date="2018-09-07T13:21:00Z", and w:initials="M". This is followed by a <w:p> element with various attributes, a <w:pPr> element containing <w:pStyle w:val="Listenabsatz"/>, <w:tabs> with <w:tab w:val="left" w:pos="579"/>, <w:spacing w:before="1" w:line="276" w:lineRule="auto"/>, <w:ind w:left="0" w:right="786" w:firstLine="0"/>, <w:jc w:val="right"/>, <w:rPr> with <w:rFonts w:ascii="Cambria" w:hAnsi="Cambria"/>, <w:color w:val="003D58"/>, and <w:sz w:val="18"/>. The code ends with </w:rPr>, </w:pPr>, <w:bookmarkStart w:id="9" w:name="_GoBack"/>, <w:r>, <w:rPr>, <w:rStyle w:val="Kommentarzeichen"/>, </w:rPr>, <w:annotationRef/>, </w:r>, and <w:r>.
- XPath/XQuery Baumeister:** A panel on the right for defining XPath or XQuery expressions.

Demo: combining XML and RDF processing

What are the advantages of such an approach?

- The document is created without an API.
- No security problems.
- Future-proof for decades as long as Docx is supported.
- Batch-compatible.
- Can be used for thousands of documents.

Demo: combining XML and RDF processing

What other formats can we integrate in this way?

- Excel
- PowerPoint
- InDesign
- SVG
- EPub
- HTML
- ...

The same technologies can also be used to transform DocX into XML, for example, and then move in the direction of RDF.

Demo: combining XML and RDF processing

If you want to test these techniques, you are welcome to try our free online tool data2check.

www.data2check.de

Agenda

- Goals of our project
- Project partners and their motivations
- Wolters Kluwer use case
- Cornelsen use case
- Demo: combining XML and RDF processing
- Conclusions and next steps

Conclusion

- Raise awareness of both Semantic Web and XML Community on mutual benefits
- Challenges for all partners are very similar, although they do different business. Therefore we think that we have a huge scalability potential here
- Existing solutions in both worlds can already help a lot, unfortunately people just do not know enough of each other (low hanging fruit)
- Knowledge engineering and explicit knowledge representation in machine readable form are key for smart content applications. XML and RDF/SKOS are perfect formats to represent this core assets
- We will work on these issues first half of 2019 and report back our results at next SEMANTiCS. Anybody interested? Please talk to us

Contact

Felix Sasaki

Cornelsen Verlag, Publishing Operations

Content Architect

Tel: +49 30 897 85-8398

Email: felix.sasaki@cornelsen.de

Manuel Montero Pineda

data2type GmbH

Managing Director, data2type GmbH

Tel: +49 6221 7391264

Email: montero@data2type.de

Christian Dirschl

Chief Content Architect

Innovation & UX

Tel: +49 173 93 15 655

Email: Christian.Dirschl@wolterskluwer.com

Cornelsen Verlag GmbH

Mecklenburgische Straße 53

14197 Berlin

cornelsen.de

data2type GmbH

Wieblinger Weg 92a

69123 Heidelberg

data2type.de

Wolters Kluwer Deutschland GmbH

Freisinger Strasse 3

85716 Unterschleißheim

wolterskluwer.com