

Von der Typographie zu inhaltlichem XML-Markup.

Automatische
Auszeichnungsverfahren am
Beispiel des Mittellateinischen
Wörterbuchs.

Referent: B.A. Alexander Trauth

Das Kompetenzzentrum Trier

- Internationale Forschungs- und Serviceeinrichtung der Sprach-, Literatur- und Medienwissenschaften mit der primären Aufgabe, digitale Publikationsformen vornehmlich in den Geisteswissenschaften zu unterstützen
- Etablierung von Technologien aus dem Bereich der Informatik in den Geistes- und Kulturwissenschaften
- Beteiligung an neubegonnenen Initiativen im Bereich des Ausbaus von Wissensverbänden und des verteilten Wissensmanagements

Das Kompetenzzentrum Trier

Leistungsspektrum

- Volltexterfassung älterer Publikationen im Double-Keying-Verfahren
- Auszeichnung der Daten in SGML/XML
- Entwicklung grafischer Benutzerschnittstellen für Offline- und Online-Publikationen

Das Mittellateinsche Wörterbuch

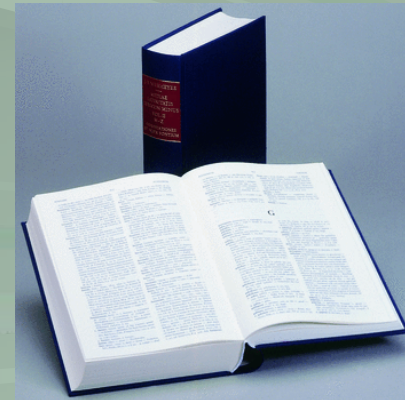
Was bedeutet Mittellatein?

- Als Mittellatein wird die lateinische Sprache des Mittelalters (ca. 500-1500 n. Chr.) bezeichnet
- Latein lebte nicht nur in den romanischen Sprachen, sondern auch als Sprache der Literatur und Kultur, der Wissenschaft, des Rechts und der Kirche weiter und wurde stets als Schulfach unterrichtet.
- Kontinuierliche Adaption an die sich wandelnde Welt

Das Mittellateinische Wörterbuch

Gegenstand

- Die lateinische Literatur des Mittelalters im deutschen Sprachgebiet
- wichtige Texte zur deutschen Geschichte, vor allem Quellen aus dem Frankenreich und aus dem sog. Reichsitalien
- fachliterarische Texte



Aufgaben der XML-Auszeichnung

Aller Anfang ist schwer

- Transkriptionen der Buchseiten ins RTF-Format bilden das Ausgangsmaterial
- Vorcodierung: unkonventionelle Auszeichnungen durch die Transkriptoren (z.B. <-P>, <zimuduan_kaishi/>, <zuo>)
- Nur formale Textmerkmale wurden ausgezeichnet, keine Sinnstrukturen

Aufgaben der XML-Auszeichnung

Methodisches Vorgehen

- Import der RTF-Daten ins Tustepformat
- Vom Großen ins Kleine: Grobe Mark-Ups dienen als Gerüst und schließlich der Simplifizierung detaillierter Auszeichnungen



Aufgaben der XML-Auszeichnung

Vier wesentliche Aufgaben der XML-Auszeichnung

1. Artikel auszeichnen
2. Lemmata und Hierarchiestufen auszeichnen
3. Bedeutungsangaben auszeichnen
4. Grammatikalische Ergänzungen auszeichnen
5. Quellenangaben auszeichnen
6. XML-Transport

Vielen Dank für Ihre Aufmerksamkeit!