

Documentverwerking (E018921)

Cursusomvang *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

Studiepunten 4.0 **Studietijd 120 u** **Contacturen** 30.0u

Aanbodsessies en werkvormen in academiejaar 2022-2023

A (semester 1)	Engels	Gent	hoorcollege	15.0u
			zelfstandig werk	15.0u

Lesgevers in academiejaar 2022-2023

Bronselaer, Antoon	TW07	Verantwoordelijk lesgever
--------------------	------	---------------------------

Aangeboden in onderstaande opleidingen in 2022-2023

	stptn	aanbodsessie
Master of Science in Computer Science Engineering	4	A
Master of Science in de informatica	4	A
Uitwisselingsprogramma informatica (niveau master)	4	A

Onderwijstalen

Engels

Trefwoorden

Document, documentverwerking, information retrieval, algoritme van Knuth, XML, XPath

Situering

Vandaag is een zeer belangrijk deel van de kennis en informatie van organisaties opgeslagen in weinig gestructureerde digitale documenten. Documentverwerking en documentenbeheer beogen het invoeren van methoden om documenten te structureren, en (soms heel grote) collecties of verzamelingen van documenten op een efficiënte manier te beheren. Hiervoor is niet alleen een elementaire kennis van tekstverwerking en documentverwerking nodig, maar ook een basiskennis van de technieken die aangewend worden voor documentopslag en terugwinning.

Inhoud

- Modellen en transformaties van documenten: Documentformaten, Overgangen van logische structuur naar fysische voorstelling, Overgangen van fysische voorstelling naar logische structuur.
- Tekstverwerking binnen documenten: Letterstijlen en lettersoorten, opsplitsen van paragrafen in lijnen via het algoritme van Knuth, bladzijdebeschrijvingstalen en het portable document format.
- XML technologie: XML (eXtensible Markup Language), de schema talen DTD, XML schema en RelaxNG, formele documentmodellen, XML transformaties.
- Information retrieval: Booleaans zoeken, index constructie, evaluatie van zoekmachines, vectorruimte-model, probabilistisch documentmodel.
- Zoeken in documenten: reguliere expressies en XPath.
- Web search: webcrawlers, Link-analyse, XML retrieval.

Begincompetenties

Basiskennis programmeren

Eindcompetenties

- 1 Kennis van de verschillende formaten voor documentopslag.
- 2 Kennis van de verschillende aspecten die horen bij de vormgeving van documenten.
- 3 Kennis van de verschillende aspecten van XML technologie en deze kunnen toepassen.

4 Kennis van de principes van "Information Retrieval" en deze kunnen toepassen.

5 Kennis van elementaire zoekmethoden in documenten

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk na gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Hoorcollege, Zelfstandig werk

Leermateriaal

De cursus nota's worden beschikbaar gemaakt op de elektronische leeromgeving, naarmate de lessen vorderen.

Referenties

- Digital Typography, Donald Knuth, CSLI Publications, 1999
- Digital Typography, An Introduction to Type and Composition for Computer System design, Richard Rubinstein, Addison-Wesley, 1988
- The Concise SGML Companion, Neil Bradley, Addison-Wesley, 1996
- The XML Companion, Neil Bradley, Addison-Wesley, 1998
- The XML Schema Companion, Neil Bradley, Addison-Wesley, 2003
- XSL Formatting Objects, Sharon Adler Ed., Sams Publishing, 2003
- Document Warehousing and text Mining, Dan Sullivan, Wiley, 2001
- Introduction to Information Retrieval, C. D. Manning, P. Raghavan, H. Schuetze, Cambridge, 2008.
- Understanding Search Engines, Michael Berry and Murray Browne, SIAM, 2005

Vakinhoudelijke studiebegeleiding

Interactieve ondersteuning via de elektronische leeromgeving (een cursusforum; studenten kunnen zelf nieuwe onderwerpen opstarten); op afspraak, na aanvraag per e-mail, voor persoonlijke aangelegenheden.

Evaluatiemomenten

periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijk examen, Openboekexamen

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijk examen, Openboekexamen

Evaluatievormen bij niet-periodegebonden evaluatie

Tweede examenkans in geval van niet-periodegebonden evaluatie

Niet van toepassing

Toelichtingen bij de evaluatievormen

De periodegebonden evaluatie bestaat uit een schriftelijk examen met open boek

Eindscoreberekening