HÖGSKOLAN I BORÅS

Textkodning av kulturarvet
Text encoding the cultural heritage
7,5 högskolepoäng
7,5 credits

Ladokkod: ÖTTK13
Version: 4.0
Fastställd av: Utbildningsutskottet 2015-02-25
Gäller från: VT 2015

Nivå: Avancerad nivå
Huvudområde (successiv fördjupning): Biblioteks- och informationsvetenskap (A1N)
Utbildningsområde: Övrigt
Ämnesgrupp: Turism- och fritidsvetenskap
Förkunskapskrav: Kursen Teknik för digitala bibliotek (ÖTDBS3), eller motsvarande, med godkänt resultat, eller dokumenterade kunskaper och färdigheter i XHTML, CSS och XML i motsvarande grad.
Betygsskala: Underkänd, Godkänd eller Väl godkänd

Innehåll

- Digitalisering av text
- textfångst och transkription
- textkodning
- eXtensible Markup Language (XML) och relaterade tekniker
- TEI (Text Encoding Initiative)
- öppen mjukvara för textkodning och XML-relaterade tekniker
- textkvalitet och filkvalitet
- lagring och arkivering
- webbpublicering och förmedling
- metadata
- Strategier för digitalisering

Mål
Efter genomgången kurs ska studenten:

med avseende på *kunskap och förståelse*
- Förklara hur XML-kodning av texter kan integreras med olika tekniker och standarder inom digitaliserade kollektioner för att stödja lagring, hållbarhet, optimering, flexibilitet och återanvändning för grupper som exempelvis humanistiska forskare
- Redogöra för tekniker för att manuellt eller maskinellt transkribera texten ur ett bilddigitaliserat material
- Redogöra för standardiserade och internationellt rekommenderade strategier, tekniker och verktyg för digitalisering och kodning av text

med avseende på *färdigheter och förmågor*
- Märka upp en digital text med relevant XML-applikation, metadata och annan bibliografisk information på ett sätt som kan anses adekvat i relation till materialets karaktär och tänkta användning
- Publicera digitaliserat och textkodat material på webben, försett med adekvata metadata och förklarande kringtexter

med avseende på *värdering och förhållningssätt*
- Kritiskt värdera dessa strategier, tekniker och verktyg och identifiera deras respektive för- och nackdelar i relation till digitaliserings- och textkodningsprojekt av olika slag

- Förklara hur val av metoder och tekniker inom kulturarvsdigitalisering villkoras av och samspelar med det omgivande samhället, med avseende på exempelvis etik och tillgängliggörande av resultat.

Undervisningsformer

Undervisningen bedrivs i form av självstudier, inlämningsuppgifter, laborationer, föreläsningar och projektarbeten.

Undervisningen bedrivs på svenska, men undervisning på engelska kan förekomma.

Examinationsformer

Kursen examineras genom laborationer, tentamina och projektarbeten.

Studentens rättigheter och skyldigheter vid examination är enligt riktlinjer och regelverk vid Högskolan i Borås.

Kurslitteratur och övriga läromedel

Burnard, Lou, O'Brien O'Keeffe, Katherine & Unsworth. John, red. (2006). Electronic textual editing. New York: MLA/TEI. (Valda delar, = ca 60 s.).

Dahlström, Mats & Ore, Espen S. (2011). Electronic text editing in Scandinavia. Ingår i Henrikson, P. & Janss, C., red. Bausteine zur Geschichte der Edition: Skandinavische Editionen. Tübingen: Niemeyer. (= ca 20 s.).

Dappert, Angela & Markus Enders (2010). Digital Preservation Metadata Standards. Information Standards Quarterly, Spring 2010. [Tillgänglig elektroniskt]

Gunnarsson, Mikael (2007). Digitalisering av text: (XHTML), CSS, TEI, XSLT. Borås: BHS. (= 100 s.).

Kungliga Biblioteket (2005). DIGSAM: Digitalisering och dess samordning inom Kungl. biblioteket. Stockholm: Kungliga Biblioteket. (= 114 s.).

Lund, Niels D. et al. (2009). Fra samling til sampling: Digital formidling af kulturarv. Copenhagen: Multivers. (Valda delar, = ca 40 s.)

Minerva (2004). Good practices handbook. Version 1.3. Svensk översättning: (Valda delar, referensverk).

Minerva (2008). Intellectual Property Guidelines. Version 1.0. [Tillgänglig elektroniskt]

Ray, Eric T. (2003). Learning XML. 2. ed. Sebastopol: O'Reilly. (= 399 s.).

Rydberg-Cox, Jeffrey A. (2009). Digitizing Latin incunabula: challenges, methods, and possibilities. Digital Humanities Quarterly 3.1. [Tillgänglig elektroniskt]

Schreibman, Susan, Siemens, Ray & Unsworth, John, red. (2004). A Companion to digital humanities. Oxford: Blackwell. (Valda delar, = ca 73 s.).

Smith, Ray (2007). An Overview of the Tesseract OCR Engine. In: ICDAR '07 Proceedings of the Ninth International Conference on Document Analysis and Recognition, vol. 2. Washington D.C.: IEEE Computer Society. [Tillgänglig elektroniskt]

Sutherland, Kathryn & Deegan, Marilyn (2009). Transferred illusions: Digital technology and the forms of print. London: Ashgate. (Valda delar, =ca 50 s.).

Tanner, Simon (2004). Deciding whether Optical Character Recognition is feasible. London: King's College. 11 pp. [Tillgänglig elektroniskt]

TEI P5: Guidelines for Electronic Text Encoding and Interchange" (2014). Oxford: The TEI Consortium, Technical Council. [Valda delar]

van Branden, Ron, Melissa Terras & Edward Vanhoutte. TEI by Example. [Tillgänglig elektroniskt]

Studentinflytande och utvärdering

Studenternas synpunkter på kursen ska systematiskt och regelbundet inhämtas i muntlig och/eller skriftlig form. Studenterna ska delges resultatet av utvärderingen. Resultaten av utvärderingarna ska ligga till grund för kursens vidare utveckling. Se vidare Policy för kursutvärdering, Högskolan i Borås 2005-06-07, dnr 56-02-10.

Övrigt

Kursen ingår i Masterprogram i Biblioteks och informationsvetenskap, distansutbildning.