

# XML-standardointi julkishallinnossa: mahdollisuuksia ja haasteita

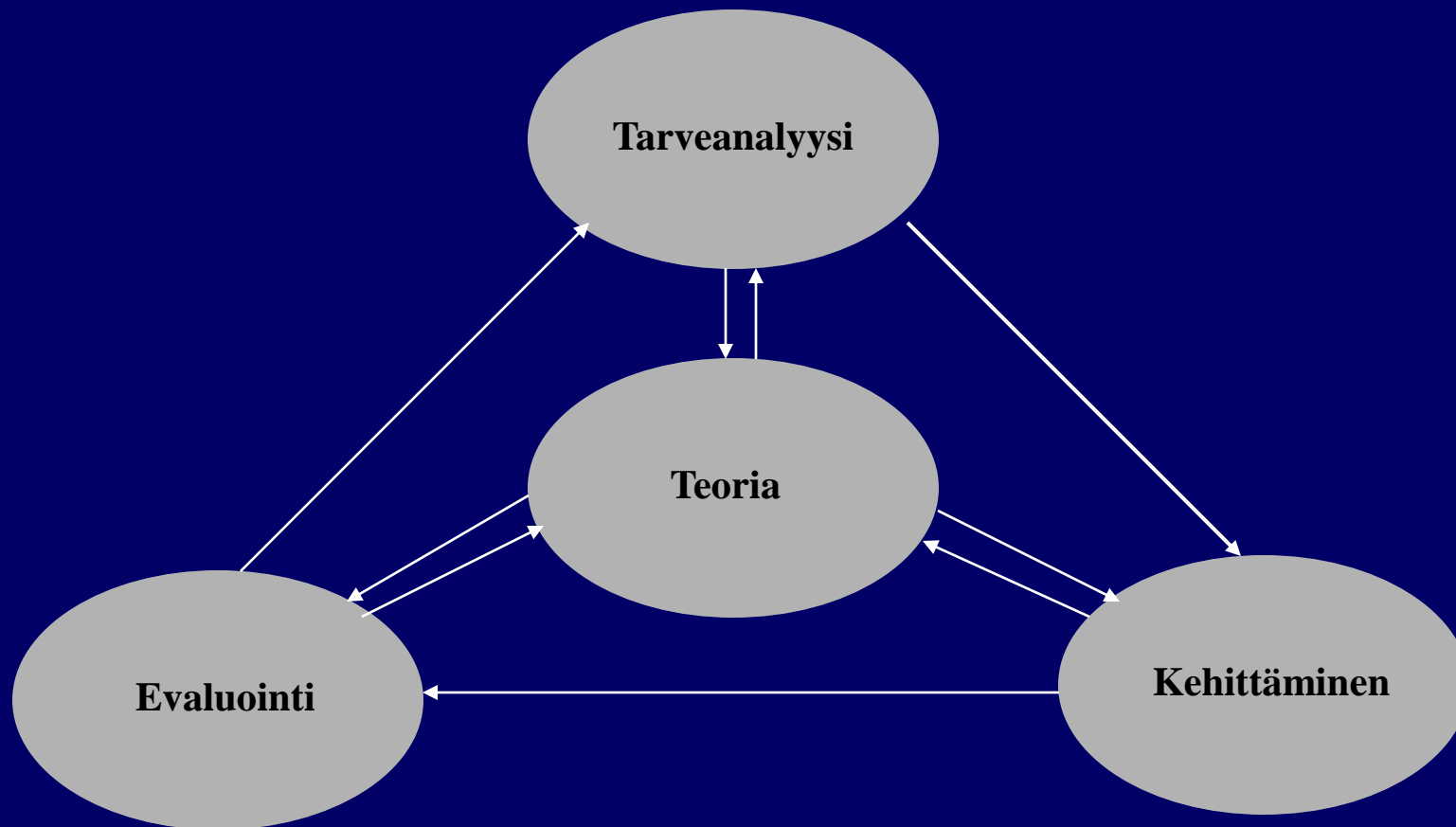
Airi Salminen  
Jyväskylän yliopisto  
<http://www.cs.jyu.fi/~airi/>

Valtioneuvoston  
tieto- ja viestintäammattilaisten  
aamukahvitilaisuus  
4.6.2003

# Sisältö

1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa
2. XML ja sen käyttö
3. XML-standardointi
4. Mahdollisuudet julkishallinnossa
5. XML-standardoinnin haasteet
6. RASKE-metodologia rakenteistamiseen
7. RASKE2-projekti
8. Yhteenveto

# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa



# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Teoria

- Mitä rakenteiset dokumentit ovat?
  - Miten niitä mallinnetaan?
- 
- Salminen, Tague-Sutcliffe & McClellan (1995)
  - Salminen & Tompa (1999); Salminen & Watters (1992)

# **1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa**

**Tarveanalyysiä, kehittämistä ja evaluointia  
yhteistyössä julkishallinnon organisaatioiden ja  
teollisuusyritysten kanssa.**

# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Julkishallinnon projekteja:

- Rakenteisten asiakirjojen kehittäminen (RASKE) 1994-1998: Eduskunta, UM, VM, Tietotehdas, Tekes, Suomen Akatemia
- European User Views to Legislative Information in Structured Form (EULEGIS) 1998-2000: EC, Eduskunta, OM, VM, Tieto, Indra, ATOS, ...
- XML and Data Warehousing in Air Defence (AirXML) 2002-2004: Ilmavoimat
- Menetelmiä julkishallinnon tietojärjestelmien ja web-palveluiden integrointiin (MetaRASKE) 2003-2006: Eduskunta, VM, OM

# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Projekteja yhteistyössä teollisuusyritysten kanssa:

- Menetelmiä SGML-standardointiin teollisuudessa (inSGML) 1998-2001: Nokia, Metso, Puolustusvoimat, Tekla, iCALS-ohjelma, Tekes
- Digital Rights of Electronic Learning Material (DrElma) 2002-2004: Kirjavälitys, Legosoft, Metso Paper, Otava, Republica, Tekes, ...

# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Tarveanalyysi

- Millaisia tarpeita ihmisillä, organisaatioilla ja yhteisöillä on rakenteisten dokumenttien hallintaan liittyen?
- Vaatimukset uusille teknologisille ratkaisuille?
  - Kiuru, Salminen & Chen (1999); Salminen & Tompa (2001)
  - Lyytikäinen, Tiitinen & Salminen (2000)
  - Tiitinen, Lyytikäinen, Päivärinta & Salminen (2000)



# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Kehittäminen

- Menetelmiä ja tekniikoita rakenteisten dokumenttien käyttöön ja käyttöönottoon.

- Karjalainen & Salminen (2000); Kuikka & Salminen (1995, 1997)
- Lyytikäinen, Tiitinen & Salminen (2000, 2001a, 2001b, 2003)
- Lyytikäinen et al. (2000); Salminen (2000, 2003)
- Salminen, Kauppinen & Lehtovaara (1997);
- Salminen, Lehtovaara & Kauppinen (1996)
- Salminen, Lyytikäinen & Tiitinen (2000); Salminen & Tompa (1993)
- Tiitinen, Lyytikäinen, Päivärinta & Salminen (2000)

# 1. Rakenteisten dokumenttien tutkimus Jyväskylän yliopistossa

## Evaluointi

- Ovatko kehitetyt ratkaisut hyviä tarkoitukseensa?
  - Mitä ongelmia ja puutteita kehitetyissä ratkaisuissa on?
  - Mitä kehitetyistä ratkaisuista voidaan oppia?
- 
- Korhonen & Salminen (2003); Salminen (1994, 1995)
  - Salminen, Lyytikäinen, Tiitinen & Mustajärvi (2000, 2001)

## 2. XML ja sen käyttö

XML = Extensible Markup Language

Joukko sääntöjä informaation esittämiseksi rakenteisina dokumentteina

T. Bray, J. Paoli, C. M. Sperberg-McQueen, & E. Maler (Eds.),  
Extensible Markup Language (XML) 1.0 (Second Edition), W3C  
Recommendation 6 October 2000, <http://www.w3.org/TR/REC-xml>

## 2. XML ja sen käyttö

### *rakenteinen dokumentti*

- rakenne määritelty ihmisen ja tietokoneen luettavissa olevalla määrittelyllä
- dokumentti koostuu nimetyistä osista
- dokumentin osia voidaan hakea ja käsitellä
- ihmiselle esittävä ulkoinen esitysmuoto mahdollista erottaa sisällöstä
- käyttö edellyttää rakenteen suunnittelua ja rakenneosien nimeämistä

## 2. XML ja sen käyttö

### *XML:n käyttötavat organisaatioissa*

- Organisaation tietovarantojen esittämiseksi, tallentamiseksi ja käsittelemiseksi standardoidussa muodossa rakenteisina dokumentteina
- Järjestelmien ja tietovarantojen yhteensovittamiseksi
  - Dokumenttien ja muiden resurssien metatietojen esittäminen ja välittäminen standardoidussa muodossa
  - Järjestelmien välinen tiedon siirto (esim. OVT-sanomavaihto)
  - Tietojen kokoaminen erilaisista järjestelmistä (esim. portaalit)

# 3. XML-standardointi

## *XML-standardoinnin tasot*

- Universaali standardointi: sääntöjä XML-kielen käytön tukemiseksi erilaisilla sovellusalueilla (esim. XML Names, XML Schema, XLink, DOM)
- Sovellusaluekohtainen standardointi: sääntöjä XML:n hyödyntämiseksi tietyllä sovellusalueella (XHTML, ebXML, RDF, XML-Signature)
- Paikallinen standardointi: säännöt ja käytännöt tietyn organisaation tai organisaatioryhmän tiedonhallintaan XML:ää hyödyntäen
  - Salminen (2003, 2000-2003)

# 3. XML-standardointi

## *XML-standardointi julkishallinnossa*

- Kansainvälinen: yhteiset säännöt tietylle kansainväliselle yhteisölle (esim. EU) tai tietylle julkishallinnon sektorille (esim. lainsäädäntö, verotus, koulutus)
- Kansallinen: esimerkiksi *Julkishallinnon XML-strategia*
- Paikallinen: esimerkiksi valtiopäiväasiakirjojen rakenteistaminen

# 4. Mahdollisuudet julkishallinnossa

## *XML:n soveltamiskohteet:*

Valtiovarainministeriö, XML-Strategiatyöryhmä, Julkishallinnon XML-strategia, 13.02.2003, <http://www.vm.fi/resource/fi/33825.pdf>

1. Yhteiset asiointiliittymät yhdistämään erilaisia perusjärjestelmiä
2. Rajapinnat perusrekistereihin perusrekistereiden ja niihin liittyvien järjestelmien väliseen tiedonvaihtoon
3. Organisaatioiden välinen tiedonvaihto (OVT/EDI)
4. Tekstitiedon rakenteistaminen
5. Monikanavajulkaiseminen
6. Tiedon haun parantaminen
7. Tiedon pitkäaikaissäilytyksen varmistaminen



# 4. Mahdollisuudet julkishallinnossa

## *Järjestelmien ja tietovarantojen yhteensovittaminen*

- Yhteiset asiointiliittymät perusjärjestelmiin
- Rajapinnat perusrekistereihin
- OVT/EDI-ratkaisut

## *Mahdollisia vaikutuksia*

- Helpottaa tiedon hakua ja asioiden käsittelyä
- Tehostaa perusrekistereiden käyttöä ja tiedon tallentumista perusrekistereihin
- Tehostaa organisaatioiden välistä tiedon vaihtoa

# 4. Mahdollisuudet julkishallinnossa

## *XML tietovarantojen tallennusmuotona*

- Edellyttää asiakirjojen rakenteistamista

## *Mahdollisia vaikutuksia*

- Helpottaa ja tehostaa tiedon hakua ja asioiden käsittelyä toimintaprosesseissa
- Tukee monikanavajulkaisemista
- Tukee tiedon pitkäaikaissäilyvyyttä
- Vähentää riippuvuutta tietystä järjestelmätoimittajasta

# 5. XML-standardoinnin haasteet

## *Paikallisen standardoinnin tasoja*

- Tekninen taso: Vaaditaan teknistä järjestelmätietämystä, palvelee järjestelmien vuorovaikutusta, ei välttämättä muutoksia tiedon tuottamistapoihin, työprosesseihin tai käyttöliittymiin. Esim. tietokantojen yhteensovittaminen.
- Semanttinen taso: Vaatii kohdealueen tietämystä.
- Sisältötaso: Vaatii asiantuntemusta sisällöstä, tuottamisprosesseista, tuottamistekniikoista, käytöstä.

# 5. XML-standardoinnin haasteet

## *Semanttinen taso*

- Vaatii sopimista keskeisistä metatiedoista, sanastoista, ontologioista.
- Vaatii yhteisymmärrystä merkityksistä.
- Vaatii tietämystä ylemmän tason standardoinnista (universaali, sovellusaluekohtainen).
- Eri tasojen ja yhteisöjen standardoinnin sykkronointi voi aiheuttaa ongelmia.
- Ratkaisuja voidaan tehdä asteittain.
  - Salminen (1994, 1995)

# 5. XML-standardoinnin haasteet

## Sisältötaso

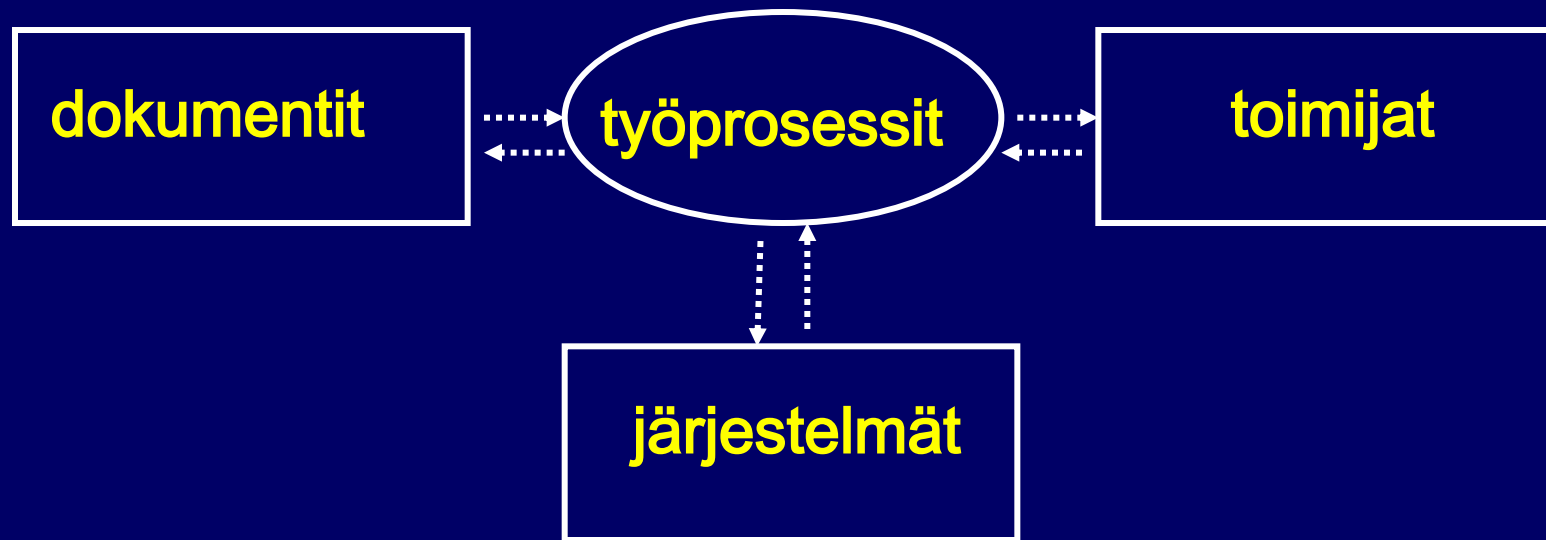
- Vaatii sopimista dokumenttistandardeista.
- Vaatii tietämystä sisältöjen tuottamisprosesseista ja ihmisten työtehtävistä.
- Tiedon tuottamistavat ja -tekniikat muuttuvat.
- Hyödyt eivät välttämättä nähtävissä välittömästi (esim. pitkäaikaissäilytys); motivointi muutokseen voi tuottaa ongelmia.
- Kehittymätön tai jatkuvasti muuttuva teknologia voi aiheuttaa ongelmia.
  - Salminen, Lyytikäinen, Tiitinen & Mustajärvi (2000, 2001)

## 6. RASKE-metodologia rakenteistamiseen

- Joukko menetelmiä ja tekniikoita kokonaisvaltaiseen dokumenttien hallinnan kehittämiseen.
- Erityisesti projekteihin, joissa tavoitteena rakenteisten dokumenttien käyttöön siirtyminen.
- Menetelmät kehitettiin sisältöjen rakenteistamiseen, ei semanttisen tason standardointiin.
- Menetelmien soveltuvuutta XML-standardointiin myös standardoinnin semanttisella tasolla tullaan testaamaan MetaRASKE-projektissa.
  - Salminen (2000, 2003), ym.

# 6. RASKE-metodologia rakenteistamiseen

## Dokumenttien hallintaympäristö



## 6. RASKE-metodologia rakenteistamiseen

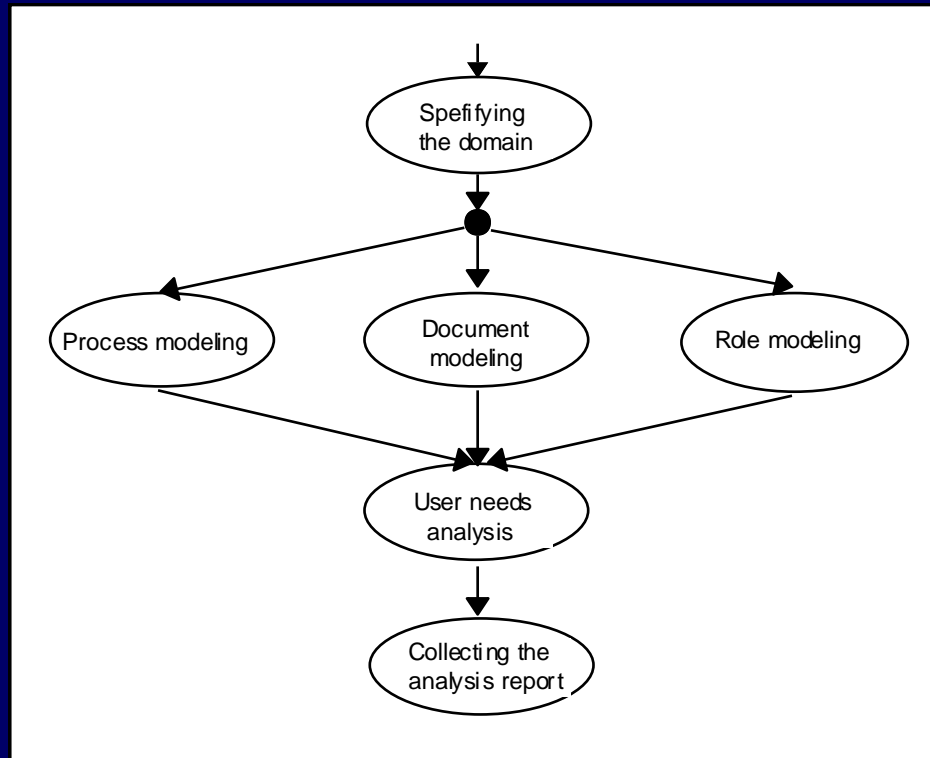
Dokumenttien hallinnan suunnittelu ja kehittäminen edellyttävät dokumenttianalyysiä. Analyysissä selvitetään ja kuvataan:

- nykyiset toimintatavat
- tarpeet tuleville toimintatavoille
- ehdotukset ja suunnitelmat tuleviksi toimintatavoiksi



# 6. RASKE-metodologia rakenteistamiseen

## Dokumenttianalyysi RASKE-metodologiassa



# 6. RASKE-metodologia rakenteistamiseen

## RASKE-metodologiassa kuvauksen kohteina ovat

- organisaation tai organisaatioverkoston toiminnot
- organisaatiot ja ihmiset toimijoina
- dokumentit ja niihin liittyvät metatiedot
- järjestelmät
- organisaatioiden ja ihmisten tarpeet

- Lyytikäinen, Päivärinta, Salminen & Tiitinen (1997)
- Tiitinen, Päivärinta, Salminen & Lyytikäinen (1997)

# 7. RASKE2-projekti

- Suunnitelma vuosille 2003-2006, alkanut 1.5.2003.
- Tavoitteena kehittää menetelmiä ja tekniikoita organisaatioverkostojen hajallaan olevien tietojärjestelmien ja web-palveluiden integrointiin metatietojen standardoinnin avulla.
- Tutkitaan kansainvälisten semanttisen webin tekniikkojen soveltamista suomalaiseseen julkishallintoon.
- Tutkimustyö kytketään suomalaisen lainsäädäntöprosessin tiedonhallinnan ja sen myötä eduskunnan ja ministeriöiden web-palveluiden parantamiseen.

# 7. RASKE2-projekti

## *Vaiheet*

- Kohdealueen analyysi: 5/2003 - 12/2003
- Standardointivaihe: 1/2004 - 12/2004
- Toteutusvaihe: 1/2005 - 12/2005
- Evaluointi- ja raportointivaihe: 1/2006 - 4/2006

## 8. Yhteenveto

- XML:llä monipuolisia käyttökohteita julkishallinnossa.
- XML-standardointia eri tasolla: universaali, sovellusalue, paikallinen. Eri tasojen standardoinnin synkronoinnin haasteet huomioitava.
- Paikallinen XML-standardointi teknisellä, semanttisella tai sisällöllisellä tasolla; näiden toteutus vaatii erilaisia valmiuksia.
- Sekä semanttisen että sisällöllisen tason standardoinnissa kohdeyhteisön ihmisten tehokas yhteistyö on keskeisin haaste.

## 8. Yhteenveto

- RASKE-projektissa kehitettiin menetelmiä rakenteistamiseen erityisesti sisältötason standardointiin.
- RASKE:n kuvauskohteina työprosessit, organisaatiot, ihmiset ja heidän työtehtävänsä, tekniikka, dokumentit, ihmisten tarpeet työtehtävissään, organisaatioiden tarpeet tavoitteidensa suorittamisessa.
- RASKE2-projektissa tullaan testaamaan, muokkaamaan ja laajentamaan RASKE-metodologian menetelmiä soveltumaan semanttisen tason standardointiin.

# 8. Yhteenveto

## Lisätietoja

- Jyväskylän yliopiston dokumenttien hallinnan tutkimusprojekteista: <http://www.cs.jyu.fi/~airi/docman.html>
- XML-kieliperheestä:  
Salminen, A., XML Family of Languages. Overview and Classification of W3C Specifications.  
<http://www.cs.jyu.fi/~airi/xmlfamily>
- Semanttisesta webistä:  
Airi Salminen, Semanttinen web: visio uudesta webistä.  
<http://www.cs.jyu.fi/~airi/presentations/SemanttinenWeb-29012003.ppt>

# Lähteitä

- Salminen, A. (2003). Document analysis methods. To appear in the 2nd Edition of the *Encyclopedia of Library and Information Science*. New York: Marcel Dekker.
- Korhonen, R. & Salminen, A. (2003). Visualization of EDI messages: Facing the problems in the use of XML. To appear in the *Proceedings of the Fifth International Conference on Electronic Commerce*. Oct. 1-3, 2003, Pittsburgh.
- Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2003). Unifying access to heterogeneous document databases through contextual metadata. In S. A. Becker (Ed.), *Effective Databases for Text & Document Management* (pp. 93-107). Hershey PA: IRM Press.
- Salminen, A. (2000-2003). XML Family of Languages. Overview and Classification of W3C Specifications. Available at <http://www.cs.jyu.fi/~airi/xmlfamily>.



# Lähteitä

- Salminen, A., & Tompa, F.W. (2001). Requirements for XML document database systems. In E.V. Munson (Ed.), *Proceedings of the ACM Symposium on Document Engineering (DocEng '01)* (pp. 85-94). New York: ACM Press.
- Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2001a). Supporting access to information created in inter-organizational processes. In A.G. Chin (Ed.), *Text Databases and Document Management: Theory and Practice* (pp. 223-241). Hersley, PA: Idea Group Publishing.
- Salminen, A., Lyytikäinen, V., Tiitinen, P., & Mustajärvi, O. (2001). Experiences of SGML standardization: The case of the Finnish legislative documents. In R.H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-Fourth Hawaii International Conference on System Sciences* (file etegv01.pdf at CD-ROM). Los Alamitos, CA: IEEE Computer Society.
- Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2001b). XML metadata for accessing heterogeneous legal databases. *Proceedings of the XML Europe 2001 Conference*.

# Lähteitä

- Salminen, A. (2000). Methodology for document analysis. In A. Kent (Ed.), *Encyclopedia of Library and Information Science*, Vol. 67 (Supplement 30) (pp. 299-320). New York: Marcel Dekker.
- Salminen, A., Lyytikäinen, V., Tiitinen, P., & Mustajärvi, O. (2000). SGML for e-Governance: The case of the Finnish Parliament. In *Proc. of the 11th International Workshop on Database and Expert Systems Applications (DEXA 2000)*, (pp. 349-353). Los Alamitos, CA: IEEE Computer Society Press.
- Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2000). Challenges for European legal information retrieval. In *Proc. of the IFIP 8.5 Working Conference on Advances in Electronic Government* (pp. 121-132). Universidad de Zaragoza.
- Tiitinen, P., Lyytikäinen, V., Päivärinta, T., & Salminen, A. (2000). User needs for electronic document management in public administration: a study of two cases. In *Proc. of ECIS 2000, European Conference on Information Systems, Volume 2* (pp. 1144-1151). Wien: Wirtschaftsuniversität Wien.

# Lähteitä

- Karjalainen, A., & Salminen, A. (2000). Bridging the gap between hard and soft information genres. In M. Khosrowpour (Ed.) *Challenges of Information Technology Management in the 21st Century, Proceedings of 2000 Information Resources Management Association International Conference*, (pp. 92-95). Hersley, PA: Idea Group Publishing.
- Lyytikäinen, V., Tiitinen, P., Salminen, A., Mercier, L., & Vidick, J.-L. (2000). Visualizing legal systems for information retrieval. In M. Khosrowpour (Ed.) *Challenges of Information Technology Management in the 21st Century, Proceedings of 2000 Information Resources Management Association International Conference*, (pp. 245-249). Hersley, PA: Idea Group Publishing.
- Lyytikäinen, V., Tiitinen, P., & Salminen, A. (2000). Graphical information models as interfaces for Web document repositories. In V. Di Gesù, S. Levialdi, & L. Tarantino, *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2000* (pp. 261-265). New York: ACM Press.

# Lähteitä

- Salminen, A., Lyytikäinen, V., & Tiitinen, P. (2000). Putting documents into their work context in document analysis. *Information Processing & Management* 36 (4), 623-641.
- Salminen, A., & Tompa, F. (1999). Grammars++ for modelling information in text. *Information Systems* 24 (1), 1-24.
- Salminen, A., Tiitinen, P., & Lyytikäinen, V. (1999). Usability evaluation of a structured document archive. In R.H. Sprague, Jr. (Ed.), *Proceedings of the Thirty-Second Hawaii International Conference on System Sciences* (file ddhfu06.pdf at CD-ROM). Los Alamitos, CA: IEEE Computer Society.
- Kiuru, J., Salminen, A., & Chen, S.-S. (1999). Digital archiving of legislative documents: A case study of Finland. Position paper in NSF Workshop on Data Archival and Information Preservation, March 26-27, 1999, Washington DC.

# Lähteitä

- **Lyytikäinen, V., Päivärinta, T., Salminen, A., Tiitinen, P. (1997).** Valtion talousarvioon liittyvien asiakirjojen rakenteistaminen. RASKE-projektin raportti. Helsinki: Eduskunta, 139 s.  
<http://www.cs.jyu.fi/~airi/raportit/RASKE-talousarvioraportti.pdf>
- **Tiitinen, P., Päivärinta, T., Salminen, A., Lyytikäinen, V. (1997).** Suomalaisten EU-lainsäädäntöasiakirjojen rakenteistaminen. RASKE-projektin raportti, Tietohallinnon selvityksiä. Helsinki: Ulkoasiainministeriö, Tietohallintolinja, 170 s.  
<http://www.cs.jyu.fi/~airi/raportit/RASKE-EU-raportti.pdf>
- **Salminen, A., Kauppinen, K., & Lehtovaara, M. (1997).** Towards a methodology for document analysis. *Journal of the American Society for Information Science* 48 (7), Special Issue on Structured Information/Standards for Document Architectures, 644-655.
- **Kuikka, E., & Salminen, A. (1997).** Two-dimensional filters for structured text. *Information Processing and Management* 33 (1), 37-54.

# Lähteitä

- Salminen, A., Lehtovaara, M., & Kauppinen, K. (1996). Standardization of digital legislative documents - a case study. In M.S. Lynn (Ed.), *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*, Vol. 5 (pp. 72-81). Los Alamitos, CA: IEEE Computer Society Press.
- Kuikka, E., & Salminen, A. (1995). Filtering structured documents in the SYNDOC environment. *Electronic Publishing - Origination, Dissemination and Design* 8 (2 & 3), 181-193.
- Salminen, A. (1995). EDIFACT for business computers: has it succeeded? *StandardView, ACM Perspectives on Standardization* 3 (1), 33-42.
- Salminen, A., Tague-Sutcliffe, J., & McClellan, C. (1995). From text to hypertext by indexing. *ACM Transactions on Information Systems* 13 (1), 69-99.

# Lähteitä

- Salminen, A. (1994). Reflections on EDIFACT: seven issues for the future international EDI message standardization. In W.R.J. Baets (Ed.), *Proceedings of the Second European Conference on Information Systems* (pp. 631-642). Breukelen: Nijenrode University Press.
- Salminen, A., & Tompa, F. (1993). PAT expressions: an algebra for text search. *Acta Linguistica Hungarica*, 41 (1-4), 277-306. Also appeared in F. Kiefer, G. Kiss, & J. Pajzs (Eds.) (1992). *Proceedings of the 2nd International Conference on Computational Lexicography, COMPLEX '92*, (pp. 309-332).
- Salminen, A., & Watters, C. (1992). A two-level structure for textual databases to support hypertext access. *Journal of the American Society for Information Science* 43 (6 ), 432-447.