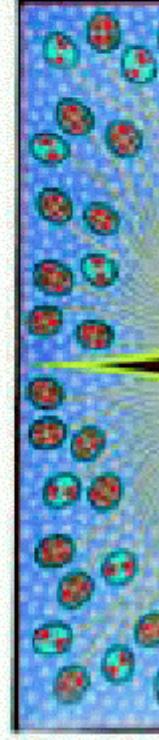

CD-ROM INCL

DEFINITIVE XML APPLICATION DEVELOPMENT



► Complete developer's guide to XML programming by a leading XML developer

Hvem jeg er

- Lars Marius Garshol
- Hovedfag informatikk på Blindern
- Jobbet et par år med XML i STEP Infotek, før jeg og noen andre startet Ontopia
- Har skrevet en XML–parser i Python, vedlikeholder 'Free XML Tools'
- Var med på å designe SAX

Boka

Fakta om boka

- Utgis av Prentice–Hall i deres Charles Goldfarb–serie
- Kommer ut i desember (håper jeg)
- Amazon sales rank: 1,756,688
- Ifølge Prentice–Hall er den på 1100 sider
- På A4–papir med 12pt Times: 400 sider
- Som enkel HTML er teksten 2 MB

Bokens del 1

- Utvikling med XML, generelle prinsipper
 - informasjonssystemer, serialisering, deserialisering
 - prosessering av XML
 - leksikalsk og logisk informasjon
 - vanlige oppgaver
 - tegnsett, hvordan håndtere dem

Bokens del 2

- Event-basert programmering
 - Forklarer hva dette er og når det bør brukes
 - Går gjennom API-ene til de ulike parserne
 - Dekker SAX meget grundig
 - Oversikt over API-et, deretter detaljene
 - Teknikker illustrert med eksempler
 - Avanserte emner som SAX-filtre dekkes også, samt SAX-parsere som ikke leser XML

Bokens del 3

- Tre-basert programmering
 - Forklarer igjen hva dette er og hva det er best til
 - DOM-API-et dekkes i detalj (a la SAX)
 - Tar med en del om hva som er galt med DOM og utilities for å hankses med noe av det
 - Dekker også alternative tre-API-er (qp_xml, Pyxie, groves, JDOM)

Bokens del 4

- XML-programmering uten programmering
 - handler om XSLT og XPath
 - relativt detaljerte innføringer gis
 - deretter vises hvordan XSLT- og XPath-motorer kan brukes i applikasjoner
 - hvordan skrive XSLT-elementer og XPath-funksjoner vises også

Bokens del 5

- Hvordan gjøre dette i Java
 - Java-utgaven av SAX dekkes
 - Det samme med DOM
 - Deretter gjennomgås XSLT-motorenes API-er (JAXP, Xalan og SAXON)

Siste del

- dekker diverse mer spesialiserte emner
 - prosesseringsmåter som ikke er nevnt hittil
 - programmering med schemas og DTD-er
 - å generere XML (fra andre datakilder)
 - to mer detaljerte eksempler

Tilleggene

- En rask introduksjon til Python for de som allerede kan å programmere
- En liten XML–ordbok
- En oversikt over XML–pakker i Python og Java, og hvordan disse installeres

Hvorfor en bok til?

- Forsøker å lære bort ting som er like gyldige om ti år, selv verktøyene da er annerledes
- Fokus er mer på prinsipper enn på detaljer
- Egentlig uavhengig av programmeringsspråk
- Gir erfaring gjennom å dekke mange ulike API-er for samme formål

Flere grunner

- Dekker en del viktige ting som ikke finnes andre steder:
 - prosesseringsmodellen og dens konsekvenser
 - ytelse
 - programmering med XSLT og XPath
 - å bruke skjemaer/DTD-er i programmer
 - tegnsett

Skrivningen

Å skrive en bok

"Writing a book is an adventure. To begin with, it is a toy and an amusement; then it becomes a mistress, and then it becomes a master, and then a tyrant. The last phase is that just as you are about to be reconciled to your servitude, you kill the monster, and fling him out to the public."

Winston Churchill

Progresjon

- **Våren '99:** Holder foredrag i Granada og Göteborg om XML-programmering
- **Høsten '99:** Foreslår boka for CFG; forslag sendes inn og aksepteres
- **Julen '99:** Skrivningen begynner, og fortsetter, og fortsetter, og fortsetter, og...

Progresjon 2

- **Påske 2001:** Teksten fullføres
- **Sommer 2001:** Copyediting
- **Høsten 2001:** Retting etter Charles' ønske
- **Nå:**
 - Bra: kun småpirk igjen
 - Dårlig: nå vet jeg hvordan jeg *burde* gjort det...

XML-programmering

Abstraksjon – hva, hvorfor, hvorfor ikke

Abstraksjon

- Å jobbe med XML er å jobbe med abstraksjoner på forskjellige nivåer
- Trikset er å velge riktig nivå for den jobben man vil gjøre
- På feil nivå blir jobben hardere, eller endatil umulig

Abstraksjon

- Ordet kommer fra latin: *abstractus*, en form av *abstrahere*, som betyr 'å trekke fra'
- En engelsk betydning er "to withdraw; to separate; to take away".
- En annen er "That which comprises or concentrates in itself the essential qualities of a larger thing or of several things."

Abstraksjonslag

- Under XML–prosessering arbeider man med flere lag av abstraksjon
- Utgangspunktet er at applikasjonen mottar en strøm av bytes
- Så skal man på en eller annen måte tolke denne strømmen
- Løsningen er å jobbe seg oppover, lag for lag

Bytes til tegn

- Det første skrittet er å gjøre om bytes til tegn
- Dvs: 3C 3F 78 6D 6C må bli til '<?xml'
- Siden XML er basert på Unicode *må* dette gjøres riktig, ellers blir resultatet bare tull
- Heldigvis er dette lett med XML: parseren gjør jobben
- *Men*, vi vet ikke lenger hva tegnkodeingen var

Tegn til tokens

- Det neste skrittet er å gjøre om
`<foo abc="def"/>`
- til
'<', 'foo', S, 'abc', '=', '"', 'def', '"', ', '>'
- Sjelden man ser på dette trinnet, men expat gjør det mulig å jobbe her (se xmlltok)

Hvorfor tokens?

- De eneste som har noe behov for å jobbe her er de som trenger informasjonen de neste trinnene kaster (editorer, stort sett)
- Ikke alle parsere gjør dette trinnet, de som ikke gjør det gir dårligere feilmeldinger
- Tapt informasjon: whitespace i markup

Events

- Dette er nivået SAX ligger på: start dokument, start element, tekst, slutt element
- Nå har vi mistet masse informasjon, men til gjengjeld er det mye lettere å arbeide
- De eneste applikasjonene som bør lages på dette nivået er de helt generelle
- Variant: pull APIs

Trær

- Dette er nivået DOM ligger på: en objektstruktur som representerer dokumentet
- XML-databaser ligger også på dette nivået
- Tapt informasjon (i forhold til SAX):
 - posisjon i opprinnelig dokument
- Noen applikasjoner med spesielle behov kan skrives her, men det er egentlig få

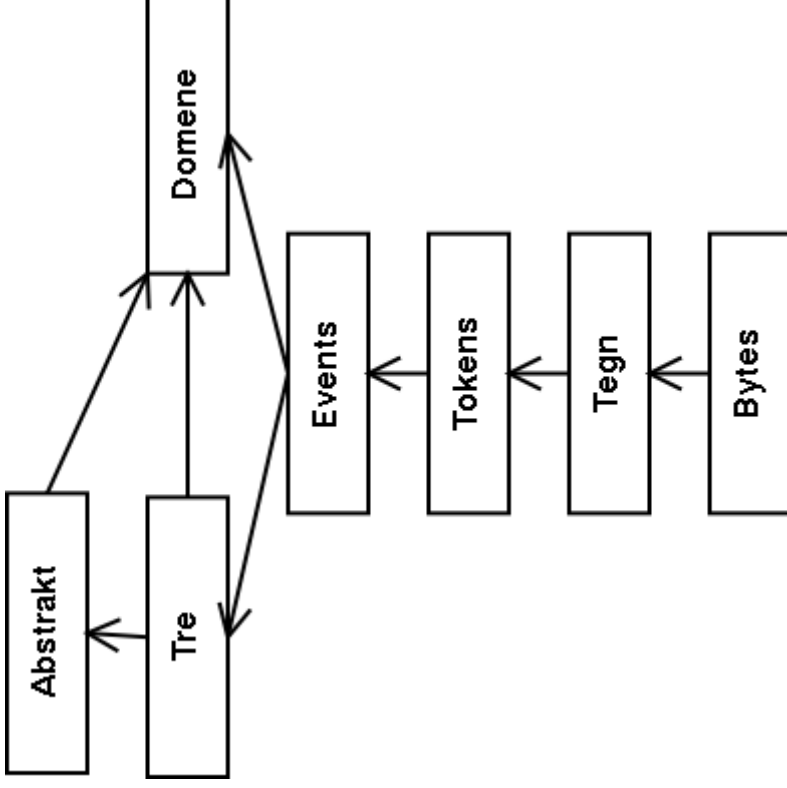
Abstrakte dokumenter

- Dette er nivået XSLT og XPath jobber på
- Her bryr vi oss ikke om objekter lenger, vi bare peker og sier "gjør!"
- Å skrive applikasjoner på dette nivået er særdeles behagelig, men ikke alltid særlig effektivt
- Tapt informasjon: noe, men ikke så mye

Domeneobjekter

- Dette er objekter som representerer dataene man ønsker å jobbe med direkte
- De må programmeres spesielt for hver eneste applikasjon man lager
- Kan en applikasjon skrives på dette nivået anbefales det på det varmeste
- Tilsvarende bruk av OR-mapping med RDBMS

Lagene



Takk for meg!

- Det var alt for denne gang
- Boken har hjemmeside på

<http://www.garshol.priv.no/download/text/ph1/>