

Kompleks data og logik i databasen
Brug af XML dokumenter til en videoapplikation

Stefan Jaensch, Jørgen Bo Arp Ladekjær

Januar 2014 - Marts 2014

Indhold

1	Introduktion	3
2	Undersøgelse af XML dokumenterne og deres struktur	4
3	Udarbejdelse af XML skema til validering af XML dokumenterne	5
4	XQuery søgning i video bibliotek	6
5	Fuld tekst søgning i video bibliotek	7
6	Join af data imellem to XML dokumenter	8
7	Brug af XML i PostgreSQL	9
8	Konklusion	10
A	Installations guide	11

Kapitel 1

Introduktion

Videoapplikationer som Netflix, Youbio og Viaplay er blevet meget populære i de seneste par år. Dette projekt handler om at skabe nogle søgefunktioner, som der må være brug for i sådanne videoapplikationer, hvor et stort antal videoer er tilgængelige for brugeren. Disse søgefunktioner skal gøre det muligt for brugeren at navigere rundt i de mange videoer og finde netop det indhold eller den specifikke video, som måtte have interesse for brugeren.

Som datakilde til dette projekt er der valgt et åbent API (Application Programming Interface) fra Danmarks Radio, som stiller stort set alt deres egenproducerede indhold tilgængeligt for brugeren/programmøren via adressen <http://www.dr.dk/nu/api>. Dette API giver fx mulighed for at hente en total oversigt over alle tilgængelige videoer, i et enkelt XML dokument, samt hente på detaljer omkring specifikke videoer.

I dette projekt hentes disse XML dokumenter ud via en standard webbrowser og gemmes lokalt, hvorefter de indlæses i en BaseX XML database, hvorfra der vil blive udarbejdet forskellige søgefunktioner til forespørgsler i XML dokumenterne.

I øjeblikket findes der desværre ingen tilgængelig beskrivelse af struktur eller elementer i de tilgængelige XML dokumenter, så derfor starter projektet med en analyse af netop strukturen eller elementerne i de XML dokumenter som er udvalgt til dette projekt.

Kapitel 2

Undersøgelse af XML dokumenterne og deres struktur

TODO: Insert stuff about study xml documents

- Description of the selected XML documents

- What documents we will use for this project?

- What is the structure of the XML document?

- What elements are there and what is described in the element?

- Proposals for changes in the structure or elements? (Is there anything that can be improved) (Attributes might be used instead of some of the elements)

Kapitel 3

Udarbejdelse af XML skema til validering af XML dokumenterne

TODO: Insert stuff about xml schemas

Description of how we create them and the choice of types.

XML validation testing.

Kapitel 4

XQuery søgning i video bibliotek

TODO: Insert stuff about xquery search

- Description of each search and how it is solved

- Find all kinds of labels for programs. Find the number of programs for each label.

- Find all kinds of broadcasting channels. Find the number of programs for each channel.

- Find programs between specific date intervals.

- Find all videos from a particular series of programs sorted by date.

Kapitel 5

Fuld tekst søgning i video bibliotek

TODO: Insert stuff about full-text search

- Using Full-text search in BaseX

- Find relevant videos based on full text search.

- Sorting results by relevance (score).

- Creating and using stop-word list

- Creating af word cloud (image before stop-word list and after stop-word list)

Kapitel 6

Join af data imellem to XML dokumenter

TODO: Insert stuff about joining data

Which XML documents to be joined and why?

Compare the same XML document different dates. (Is something added,removed,updated
? list results)

Kapitel 7

Brug af XML i PostgreSQL

TODO: Insert stuff about xml in postgresql

- Importing xml into PostgreSQL

- Adding full-text search

- Compare full text performance between BaseX and PostgreSQL

Kapitel 8

Konklusion

chapter:conclusion TODO: Insert stuff about conclusion

Bilag A

Installations guide

TODO: Insert stuff about installation guide