

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

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

This mini-project is about creating various backend functions for a video application. These backend functions shall make it possible for the user to navigate and search for specific videos in a large video directory using the BaseX XML database.

The XML documents describing all the videos in this video directory is delivered by the Danish national broadcaster Danmarks Radio, who have made their videos available through an open API using XML.

All XML documents used in this project is downloaded from this site: www.dr.dk/nu/api via a standard webbrowser and afterwards loaded into the BaseX XML database for further processing.

There is no description available for this open API, there for the first task in this project is to study and describe the content and structure of the available XML documents.

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