



## XHTML - Hvad 'De Skriftkloge' ikke fortæller dig - (Del 2)

**Dette er ikke en tutorial eller generel, enkeltstående artikel om XHTML. Forudsætningen for at få fuldt udbytte af artiklen er, at du har læst den første artikel i serien.**

Skrevet den **03. Feb 2009** af **olebole** | kategorien **Programmering / (D)HTML** | ★★★★★

### XHTML - overgangen er problemet

Da der er temmelig stor forskel på HTML og ægte XHTML, valgte man at 'liste' teknologien ind via et par versioner. Først XHTML2.0 bliver 'ægte' XHTML - altså ren XML uden HTML-understøttelse - og i denne overgangsfase ligger kimen til alle misforståelserne.

Hele idéen med XHTML er som sagt, at koden skal parses af klientens XML-parser. I XHTML version 1.0 kan koden dog både parses som HTML og XML - og i 1.1 burde den kun kunne parses som XML, men parses også som HTML.

Parses koden som HTML, har man ingen fordele ved XHTML fremfor velskrevet HTML4.01. Parses koden derimod som XML, går det væsentligt hurtigere - og fordelene ved XHTML slår igennem. Parses XHTML-kode som HTML, vil HTML-parseren endda opfatte koden som fejlfyldt, da der i HTML ikke er mulighed for at lukke åbne tags som ``

Et problem er, at en skræmmende stor del af det XHTML, der bliver skrevet, aldrig bliver parsed som XML ... også selvom koden er hyper-valid XHTML1.0-Strict eller XHTML1.1-Strict (mere om, hvorfor senere). Det vil sige, at klienten modtager et dokument med en DTD, der fortæller, at der er tale om et XHTML-dokument - på trods af, det ikke kan parses som sådan -> første fejl, der skal repareres! Derefter kommer så alle de lukkede tags, der også skal repareres, før dokumentet endelig kan parses.

Kan dokumentet ikke parses som XHTML, ville browserne altså være bedre stillet, om der var skrevet valid HTML4.01 med den korrekte DTD ;o)

### XHTML er andet og mere end lukkede tags

Alle dokumenter har tilknyttet en MIME-type, der fortæller computeren, hvad den skal gøre med det pågældende dokument. Det gælder både lokale filer og webfiler på en server.

Når dokumenter distribueres over nettet, sætter serveren denne MIME-type i dokumentets HTTP-header - i feltet 'Content-type' - og denne MIME fortæller altså klienten, hvad det er for en type dokument, så den kan vide, hvad den skal stille op med det.

Bliver klienten mødt af en MIME-type, den ikke understøtter, vil den normalt åbne en dialog, hvor brugeren bliver tilbudt at downloade dokumentet.

For HTML dokumenter er MIME-typen 'text/html'.

For XHTML dokumenter 'application/xhtml+xml'.

Eller rettere:

For XHTML dokumenter burde MIME-typen være 'application/xhtml+xml'.

Der hersker i øjeblikket et syndigt rod i MIME-types for XHTML - og browsernes understøttelse af dem. Som hovedregel skal typen være 'application/xhtml+xml', hvis XML-parseren skal tage over. Sendes typen 'text/html', vil dokumentet altid blive parsed som HTML4.01 - og ikke som XML.

Har man ikke selv foretaget sig noget aktivt for det, vil XHTML dokumenter normalt blive serveret som 'text/html' - og altså mangle XHTML'ens fordele.

Er der tale om dynamiske dokumenter, er det temmelig let at sætte den korrekte MIME-type. I PHP kan det f.eks. gøres med:

```
Header("Content-type: application/xhtml+xml");
```

Er der tale om statiske dokumenter, må man ind og pille i serveren for at få den til at sende dokumenterne med korrekt MIME. Her er der forskellige muligheder, hvis beskrivelse ikke falder ind under ambitionerne for denne artikel.

Endnu et problem er, at IE6 endnu ikke understøtter 'application/xhtml+xml' <a href="<http://www.w3.org/People/mimasa/test/xhtml/media-types/results>">[ Browser Oversigt ]</a> så, når der f.eks. er tale om en IE, må man altså servere dokumentet med en anden MIME-type. Her er 'panik-løsningen' at anvende enten 'application/xml' eller 'text/xml'. To XML-typer, der begge tillades som nødløsninger, hvis man (i IE) bruger flg. 'hack':

```
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet type="text/xsl" href="copy.xsl"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
... osv, osv.
```

Stylesheet'et 'copy.xsl' skal så indeholde:

```
<stylesheet version="1.0" xmlns="http://www.w3.org/1999/XSL/Transform">
  <template match="/">
    <copy-of select="."/>
  </template>
</stylesheet>
```

Det har bare den bagdel, at når der i koden står noget før DTD'en, går IE i quirks-mode og understøtter således ikke længere CSS-standarden ... ikke mindst box-modellen går det helt galt med.

Helt generelt gælder, at hvis kode skal parses som XML, skal XML-prologen:

```
<?xml version="1.0" encoding="iso-8859-1"?>
```

stå før DTD'en. Så kan IE som sagt ikke længere tolke CSS korrekt, hvorfor der skal altså serveres speciel CSS til den.

XHTML1.1 er i modsætning til 1.0 ren XML og må som sådan slet ikke serveres som 'text/html' - men kun med en XML-relateret MIME-type. Koden parses dog alligevel som HTML, hvis den bliver sendt som 'text/html' ... hvem sagde, XHTML er enkel og logisk at gå til her i overgangsfasen? :o)

Fra XHTML2.0 er det udelukkende MIME-typen 'application/xhtml+xml', der kan bruges. Så er overgangsfænomenerne forbi og dokumentet parses som ren XML - og kan ikke andet. Til gengæld skal koden holde sig helt strengt til standarderne. Ellers 'vælder' browseren simpelthen og viser en XML-fejl ... så er det endelig slut med sloppy kode ;o)

Du kan selv checke dels dine egne sider og dels kan du jo muntre dig med at se, hvor få sider, du kan finde, der kan parses som XHTML med W3C's udvidede validator-interface: <a

href="<http://validator.w3.org/detailed.html>">[ HER ]</a>

Husk, at afkrydse 'Verbose Output', så kan du se, hvad serveren sætter som MIME-type.

### XHTML er meget mere end lukkede tags

Udover MIME-problematikken skal der også scripts anderledes i XHTML, end der skal i HTML - også scripts har afgørende betydning for, hvordan kode parses.

Det kikker vi på i endnu en artikel, hvor jeg også vil gennemgå lidt af de spændende nyheder, der venter på dørrintrinnet.

### Foreløbig konklusion

En egentlig konklusion bestående af både en samlet, faktuel opstilling og en velbegrunnet, personlig vurdering, vil jeg stadig vente med, til serien er helt færdig.

Det er dog svært at skjule, at jeg personligt er en meget varm tilhænger af XML/XHTML og glæder mig som et lille barn til Juleaften til den dag, hvor XHTML2.0 er implementeret i de vigtige browsere.

Indtil da anser jeg det til gengæld for et umådeligt spild af tid, at skrive XHTML. I mellemtiden prøver jeg i stedet at strømline min HTML4.01 kode henimod noget, der kun på ganske få parametre adskiller sig fra XHTML - og det samme gælder min scripting.

På den måde kan jeg skifte til XHTML, overnatt - samtidig med, at min kode løbende er valid og så god som den kan blive, set fra browserne ... bedre end XHTML ville være.

Dette er for det første en helt personlig vurdering, som man er mere end velkommen til at være uenig i - og for det andet, kommer der som sagt en mere nuanceret, uddybende og begrundet konklusion på et senere tidspunkt.

-----

Et lille PS:

Brugeren frol har angivet et link til en artikel, der omhandler, hvordan man kan serve sine dokumenter med den korrekte MIME - men dette løser desværre kun 'et hjørne' af problematikken.

For det første bør løsningen kun anvendes, hvis der er virkelig god grund til at anvende XHTML (= hvis det er tvingende nødvendigt). Det er nemlig en temmelig servertung løsning, der frådser kraftigt med serverens resourcer.

Dernæst skal man være opmærksom på, at hvis man bruger scripts, skal disse være skrevet til brug under XML. Almindelig scripting, som det bruges under HTML, kan ikke umiddelbart anvendes under XHTML - der har nogle andre script-DOM bindings.

Det vil der komme mere om i den næste artikel, som jeg desværre endnu ikke har haft tid til at få skrevet færdig .... hav venligst lidt tålmodighed :)

En par kommentarer til brugeren 'tumblehund':

Problemet er jo netop, at ens (java)scripts **ikke** nødvendigvis er kompatible med XHTML, når version 2.0 bliver en realitet.

Idag spænder browserne sikkerhedsnet ud under kodere, der ikke kan scripte op mod XML-DOM'en - og lader i stedet HTML-parseren overtage dokumentet. Når XHTML2.0 kommer, sker dette ikke længere - og langt de fleste XHTML-sider, der idag ligger på WWW, vil gå ned med en XML-fejl ... hele Eksperten inkluderet ;o)

- og jo, jeg har testet ... både på frontend og backend. Ellers ville der jo blot være tale om hovedløs påståelighed. Uden at jeg i den anledning ønsker at pege fingre, kan jeg ikke undlade at tvivle på, du selv har testet noget somhelst omkring performance og dårlig XHTML-kode kontra god HTML-kode - samt hvad der egentlig skal trækkes af last på serveren ;o)

Til læsere, der finder artiklen er 'farvet':

Hvis græsset er grønt, hvorfor skal jeg så tvinges til at fotografere det i sort/hvidt?

Fordi en vis procentdel af befolkningen er farveblind, skal de, for hvem farverne ikke er skjult (eller fremstår forvrænget), vel ikke luske rundt i dunkle kældre og tale om alle farverne i det skjulte. Vi skal heller ikke tvinges til at omtale græsset som 'rødt' fire gange, fordi vi har sagt 'grønt' fire gange. Den slags demokratisme er i al respekt ikke noget, jeg agter at ligge under for. Take it or leave it!

Til læsere, der måtte undre sig over brugeren stich's klynkerier, må jeg nøjes med at henvise til en lang række tråde om emnet - afsluttende med denne:

<http://www.eksperten.dk/spm/711580>

Jeg orker ikke at replicere yderligere!

#### **Kommentar af erikjacobsen d. 17. Feb 2005 | 1**

Og andre mener det samme: <http://www.sitepoint.com/newsletter/viewissue.php?id=3&issue=107#4>

#### **Kommentar af mcookie d. 15. Jan 2006 | 2**

Læs bare hvad de andre har skrevet ;)

#### **Kommentar af simonvalter d. 07. Dec 2004 | 3**

+1

#### **Kommentar af ahave d. 11. Jun 2005 | 4**

Super artikel, kan klart anbefales.  
/ahave

#### **Kommentar af milo d. 08. Dec 2004 | 5**

Yderst brugbar info :)

#### **Kommentar af nielsmo d. 07. Dec 2004 | 6**

Kan vist kun give den 13... desværre ;)

#### **Kommentar af ibtage d. 13. Dec 2004 | 7**

Læst med stor interesse for emnet

#### **Kommentar af jensgram d. 07. Dec 2004 | 8**

Ind med det - igen!

#### **Kommentar af ysubhi d. 08. Jul 2005 | 9**

super!

#### **Kommentar af tumlehund d. 13. Apr 2006 | 10**

Okay, kan vi få armene ned over denne artikel...

olebole har fat i en række problemstillinger omkring XHTML, som er svære at komme uden om. Desværre er artiklen utrolig farvet, og det gør den utroværdig. Fx:

"Parses koden som HTML, har man ingen fordele ved XHTML fremfor velskrevet HTML4.01". Det er rigtigt at man umiddelbart ikke får nogen fordel, men på det tidspunkt hvor man skal til at overgå til XHTML 2.0, er det da en kæmpe gevinst at man allerede har "XHTML-kompatibel" kode, fremfor man (igen) skal til omskrive sin kode (fra fx HTML 4.01). Desuden har man også stor fordel af at alle ens scripts allerede er "XHTML-kompatible".

Et andet eksempel: "Parses XHTML-kode som HTML, vil HTML-parseren endda opfatte koden som fejlfyldt, da der i HTML ikke er mulighed for at lukke åbne tags som ". Dette har jo ingen betydning i praksis! Jeg tvivler desuden på, at olebole har foretaget en måling.

Endnu et eksempel:

"Kan dokumentet ikke parses som XHTML, ville browserne altså være bedre stillet, om der var skrevet valid HTML4.01 med den korrekte DTD ;o)". Browseren er ikke dårligere stillet om det er XHTML "tag soup" eller HTML "tag soup" (i praksis)!

Sidste eksempel (kommentar til content negotiation):

"Det er nemlig en temmelig servertung løsning, der frådser kraftigt med serverens ressourcer.". Det er jo en overdrivelse af dimensioner! Og hvis man alligevel bruger fx ASP, PHP, ASP.NET el.lign, så betyder det altså ikke noget i det samlede billede. Desuden vil man jo, selv hvis IE engang forstår "application/xhtml+xml", skulle bruge content negotiaion så tidligere versioner får siderne serveret som "text/html".

Og husk så lige, at en XHTML-validering giver en mere præcis validering end HTML-validering, fx:

```
<p>Se min  
  <table><tr><td>fede tabel</td></tr></table>  
den er for sej</p>
```

Her vil en HTML-validering fortælle at der står et ensomt slut-p-element, mens en XHTML-validering vil fortælle at der er kommet et blokelement ind i noget der kun må indeholde inline-indhold.

Jeg vil opfordre til at bruge XHTML, hvis:

- man skal lave et nyt site
- man skal lave grundig refactoring/redesign

Hvis man allerede har et HTML 4.01 site (som validerer vel at mærke!), er der IMO ikke rigtig nogen grund til at bruge tid på at "opgradere" til XHTML, med mindre man har tid til det, eller man synes det er sjovt/interessant.

Det er under ingen omstændigheder et argument, at XHTML parses for langsomt som "text/html" (det har ingen betydning i praksis), derimod er det en fordel at ens kode allerede er klar, når IE engang forstår "application/xhtml+xml".

#### **Kommentar af roenving d. 10. Dec 2004 | 11**

Oki, vi venter spændt på fortsættelsen !-)

#### **Kommentar af skovenborg d. 08. Dec 2004 | 12**

Tja, lidt ligesom den forrige artikel - altså stadig god, velskrevet og spændende. Kan nu godt se at også mine sider bliver parset som fejlfyldt html4.01 :-S

Om jeg er blevet omvendt fra en (måske) dårlig vane er så spørgsmålet, for jeg bliver lidt i tvivl om det virkelig bliver så let at konvertere korrekt skrevet html4 til xhtml2.0 (hvis man pludselig ikke må lukke

img-tags og den slags).

#### **Kommentar af basementjack d. 08. Dec 2004 | 13**

Yderst brugbar...

#### **Kommentar af popsy d. 07. Dec 2004 | 14**

Hmm... Touche! Godt og velskrevet, rigtigt interessant.

#### **Kommentar af phatlasse (nedlagt brugerprofil) d. 07. Dec 2004 | 15**

Mindst lige så god som den første - virkelig lækkert for øjnene :-)

#### **Kommentar af 3xm d. 07. Dec 2004 | 16**

haps!

#### **Kommentar af mm12010 d. 12. Dec 2004 | 17**

#### **Kommentar af webcreator d. 08. Jul 2005 | 18**

Fantastisk - fabelagtigt. Endnu en god artikel. Jeg var faktisk ved at omskrive mine HTML 4.01 Strict sider til XHTML, men har efter gennemlæsning af denne artikel besluttet mig for, ligesom Olebole, blot at forberede koden til det. Jeg ser meget frem til den næste artikel - skynd dig, Olebole :-)

#### **Kommentar af googolplex d. 09. Dec 2004 | 19**

#### **Kommentar af stich d. 26. May 2006 | 20**

olebole omtaler XML-prologen -- og viser herefter XML-deklarationen (som kun er en del af XML-prologen). Og XML-deklarationen er altså meget sjældent påkrævet, den er blot anbefalet - og kan trygt udelades til IE.

Og IE6 forstår altså ikke application/xhtml+xml (i modsætning til hvad der står i denne rant), ej heller vil IE7 komme til det.

Men man kan bare nøjes med at sende XHTML1.0-sider med den anbefalede application/xhtml+xml mimetype til UAs der accepterer den, og som text/html til resten (dvs. bruge en del af det script frof henviser til); det er på ingen måde noget der kræver særligt meget af serveren.

Artiklen har efter min mening alt for mange farvede ord og sætninger, fx "hyper-valid" og "Koden parses dog alligevel som HTML, hvis den bliver sendt som 'text/html' ... hvem sagde, XHTML er enkel og logisk at gå til her i overgangsfasen? :o|" - at XHTML kan blive parsed som HTML har for så vidt intet med XHTML at gøre, det kan ethvert dokument (med variende succes...).

---

olebole går som sædvanlig over til at kommunikere på en alt andet end sober og anstændig måde, når andre folk kritiserer hans messen. Det gider jeg ikke indlede mig på, men vil blot lade det tale for sig selv...

Til gengæld vil jeg opfordre til at læse andre synspunkter på sagen, fx

<http://kurafire.net/articles/case-for-xhtml> og <http://www.w3.org/MarkUp/2004/xhtml1-faq>

#### **Kommentar af jpvj d. 07. Dec 2004 | 21**

Ole: hvad kan jeg sige... hvad jeg forventede fra dig!

### **Kommentar af zyph d. 10. Dec 2004 | 22**

Lækkert, rigtig godt skrevet, og det er herligt, at nogen endeligt kommer mere detaljeret ind på emnet !-)

### **Kommentar af michaeltajo d. 27. Nov 2006 | 23**

Jeg har ind i mellem følt mig stødt over din noget barske retorik; men din viden kan man ikke udsætte noget på. Artiklen er meget oplysende, afmystificerende og ikke mindst velskrevet. Parser-problematikken har jeg aldrig før skænket en tanke. Høj klasse & keep up the good work...

### **Kommentar af htmlkongen d. 08. Dec 2004 | 24**

De andre har vidst talt! :) /Htmlkongen

### **Kommentar af freakstyle d. 07. Jun 2006 | 25**

Jeg blev klogere

### **Kommentar af dxs d. 22. Dec 2004 | 26**

Rigtig godt!!

### **Kommentar af maqhem d. 31. Jan 2008 | 27**

Rigtig god artikel. Jeg glæder mig, til der kommer en 3'er. Ikke så meget som én eneste negativ kommentar fra mig, for jeg synes tydeligt, jeg kan se, når du kommer med subjektive holdninger og objektive fakta.

### **Kommentar af suxor d. 18. Dec 2004 | 28**

### **Kommentar af jonashn d. 27. Feb 2005 | 29**

Bravo!

### **Kommentar af froel d. 27. Dec 2004 | 30**

se her: [http://keystonewebsites.com/articles/mime\\_type.php](http://keystonewebsites.com/articles/mime_type.php)

### **Kommentar af wicez (nedlagt brugerprofil) d. 17. Dec 2004 | 31**

meget velskrevet artikel - godt arbejde ole

### **Kommentar af tjanum d. 26. May 2006 | 32**

Synes det er en rigtig fin artikel. Det er da rigtigt, at artiklen måske er svagt farvet, men hvad i alverden er der i vejen med at have en mening og give den til kende - specielt når den nu også indeholder en masse faktisk indhold, der sikkert er brugbar for en masse. Topkarakter giver jeg ikke, men udelukkende fordi jeg kunne have ønsket mig lidt dokumentation for dine test omkring serverbelastning. Jeg tvivler ikke på, at du har testet det, men jeg er overrasket over dine "fund", og kunne derfor godt have tænkt mig at høre om, hvordan det er testet og hvilke præcise, målbare resultater, du har fået. Men som sagt "thumbs up" herfra :-)

### **Kommentar af freehelp d. 29. Dec 2004 | 33**

### **Kommentar af morteeart d. 07. Dec 2004 | 34**

Igen et fantastisk stykke arbejde.

Ud fra dette kan man næsten drage den konklusion at man så ikke bør bruge xhtml så længe at IE går i quirks mode over "<?xml version="1.0" encoding="iso-8859-1"?>".

Eller så ikke lave design med css. (måske xslt, men det er en anden sag)

### **Kommentar af cronaldo d. 07. Jan 2006 | 35**

At underdrive: FREMRAGENDE!

### **Kommentar af jesper-moeller d. 23. Jul 2005 | 36**

Super skrevet....

BRAVO

Dejligt at nogen overskudet at udtrykke på skrift, hvad der egentlig burde råbes ud fra byens tårne.

BRAVO

Du burde skrive for nogle blade ;-)

Tjek denne side for et eksempel af en af de sjældne ÆGTE velskavet XML side (Advanced side :-)  
og læg mærke til IE advarslen...\*G\*

<http://golem.ph.utexas.edu/~distler/blog/index.shtml>

Mere god læsning om emnet XHTML XML m.m. :

<http://www.goer.org/Journal/2003/Apr/index.html>

<http://hsivonen.iki.fi/xhtml-the-point/>

<http://www.hixie.ch/advocacy/xhtml>

<http://software.hixie.ch/utilities/cgi/xhtml-for-ie/>

<http://www.damowmow.com/playground/xhtml-in-uas.xhtml>

### **Kommentar af slst d. 09. Apr 2006 | 37**