JouTo

Toimittajan työväline

Toteutusdokumentti

Versio 0.2

12.4.2001
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos
Ohjelmistotuotantoprojekti, kevät 2001, ryhmä 3

Projektiryhmä:	Petri Hellemaa (PH)
	Antti Koskinen (AK)
	Lari Laine (LL)
	Jarno Saarto (JS)
	Eija Teitto (ET)
	Harri Tuuloskoski (HT)
Ohjaaja:	Kim Viljanen
Vastuuhenkilö:	Raine Kauppinen
Asiakkaat:	Helena Ahonen-Myka
	Greger Lindén

Dokumentin versiohistoria

0.1	6.4.2001	Ensimmäinen hahmotelma
0.2	12.4.2001	Liitetty yleiskuvaus ja muutokset suunnitteluun

Sisältö

1 Johdanto (AK)

Tässä kappaleessa käydään läpi tämän dokumentin tarkoitus ja dokumentissa käytettävien määritelmien ja lyhenteiden kuvausta. Johdanto-luvusta löytyy myös viitteet sekä dokumentin rakenteen selvitys.

1.1 Dokumentin tarkoitus

Tässä toteutusdokumentissa esitellaan JouTo-ohjelmistotuotantoprojektin tuotteena syntyvän uutistoimittajan työvälineen toteutus. JouTo on lyhenne englannin kielen sanoista "A Journalist's Tool".

1.2 Määritelmät, termit, lyhenteet

CSS Cascading Style Sheets, W3C:n kehittämä, suosittu HTML- ja XML-dokumenttien ulkoasun määrittelyyn tarkoitettu kieli.
Lisätietoja: `http://www.w3.org/Style/CSS/`

DoReMi Document Management, Information Retrieval, Text and Data Mining. TKTL:n tutkimusryhmä, joka vetää TYTTI-projektia.
Lisätietoja: `http://www.cs.helsinki.fi/research/doremi/`

DTD Document Type Definition, rakenteisten dokumenttien rakenteen määrittely.

EPS Encapsulated PostScript, kehystetty versio PostScript-kielisestä ohjelmasta. Ohjelmatiedostoon lisätty kehys määrittää rajat tiedoston sisällölle ja mahdollistaa sen sijoittamisen haluttuun kohtaan sivua.

HTML Hypertext Markup Language, WWW-sivujen sisällön kuvauksessa käytetty koodaus.
Lisätietoja: `http://www.w3.org/pub/WWW/MarkUp/`

IPTC International Press Telecommunications Council, kansainvälinen yhteistyöelin, joka organisoi uutisten määrittelyä ja ryhmittelyä.
Lisätietoja: `http://www.iptc.org/`

Java Sunin kehittämä ohjelmointikieli ja järjestelmäriippumaton ajoympäristö.
Lisätietoja: `http://www.sun.com/java/`

Javadoc Sunin kehittämä java-koodin dokumentointikieli
Lisätietoja: `http://www.sun.com/java/`

Java SDK Sunin kehittämä Java-ohjelmointiympäristö.
Lisätietoja: `http://www.javasoft.com/products/`

JDBC Java Database Connectivity, Java-kielen rajapinta tietokantojen käyttöön.

NITF News Industry Text Format, toimittajien kehittämä ja standardoima XML-pohjainen uutisartikkelien määrittelykieli. Lisätietoja: `http://www.nitf.org`

PostScript Adoben kehittämä suosittu sivunkuvauskieli. Tulostettava sivu voidaan lähettää PostScript-kirjoittimelle tiiviinä PostScript-ohjelmana tilaavievän raakadatan sijaan.

TKTL Helsingin yliopiston tietojenkäsittelytieteen laitos.
Lisätietoja: `http://www.cs.helsinki.fi`

TYTTI Tietotyöläisen työasema. Vuosina 2000-2002 käynnissä oleva tutkimusprojekti Helsingin yliopiston tietojenkäsittelytieteen laitoksella, rahoittajina TEKES sekä viisi suomalaisyritystä: Nokia, Alma Media, Sanoma-WSOY, Vaisala ja Lingsoft. TYTTI-projektin vastuuhenkilö Helena Ahonen-Myka ja projektipäällikkö Greger Lindén toimivat JouTo-projektin asiakkaina.
Lisätietoja: `http://www.cs.helsinki.fi/research/doremi/TYTTI/index.html`

W3C World Wide Web Consortium, kansainvälinen yhteistyöelin, joka koordinoi WWW-tekniikan ja -standardien kehitystä.
Lisätietoja: `http://www.w3.org/pub/WWW/Consortium/`

WWW World-Wide Web, Internetissä toimiva palvelu, joka mahdollistaa tekstin ja kuvien jakamisen helposti ja edullisesti verkon lukemattomille käyttäjille. WWW:ssä aineisto jaetaan sivuihin, jotka sijoitetaan WWW-palvelimelle tavallisina tiedostoina ja hakemistoina.

XML Extensible Markup Language, W3C:n kehittämä rakenteisten dokumenttien määrittelykieli.
Lisätietoja: `http://www.w3.org/XML/`

XSL Extensible Stylesheet Language, W3C:n kehittämä XML-dokumenttien ulkoasun määritykseen tarkoitettu kieli.
Lisätietoja: `http://www.w3.org/Style/XSL/`

1.3 Viitteet

NITFDTD News Industry Text Format DTD version 2.5 8/2000,
`http://www.nitf.org`

HTTP-protokolla Hypertext Transfer Protocol – HTTP/1.1,
`ftp://ftp.isi.edu/in-notes/rfc2616.txt`

1.4 Dokumentin rakenne

2 Järjestelmän yleiskuvaus

Luvussa kuvataan järjestelmän yleinen toiminta ja sen käyttäjät sekä esitellään millaisissa ympäristössä järjestelmä on toimii.

2.1 Yleinen toiminta

Ohjelmiston arkkitehtuuri on asiakas-palvelin pohjainen. Asiakaskoneissa pyöritettävä sovellus vaatii Pentium-teholuokan koneen, jossa on toimiva tietoliikenneyhteys. Palvelimen puolen tehovaatimukset on esitetty määrittelydokumentissa kappaleessa 7.2. *Palvelimen suorituskyky.* Toimintaynpäristöksi oletetaan toimituksen sisäinen verkko.

Asiakkaan puolen ohjelmisto on WWW-selaimessa toimiva Java-sovelma (applet), ja palvelimen puolen ohjelmisto (servlet) huolehtii tietojen tallennuksesta sekä latauksesta tietokannasta.

2.2 Käyttäjät

Jouto-projektin editorin käyttäjiä ovat eri uutisalueiden toimittajat. Uutistoimittajien tehtäviin nykyisin kuuluu varsinaisen toimittajan työn lisäksi myös artikkelien kirjoitus ja muokkaus sähköiseen muotoon. Editori-osa muokkaa käyttöliittymän valitun artikkelin mukaisesti ja huolehtii että käyttäjä voi luoda vain valitun artikkelin mukaisia rakenteita. Tätä työtä varten voidaan käyttää JouTo-projektin tuotetta, jonka käyttäminen ei edellytetä suurta tietoteknistä osaamista toimittajilta. Editori-osa toimii rajapintana käyttäjän ja varsinaisen ohjelmiston toiminnan välillä.

Ohjelmalla tuotettujen artikkeleiden käyttäjäpiiriin kuuluu myös ylläpitohenkilöitä. He voivat hakea toimittajien kirjoittamia artikkeleita tietokannasta ja taittaa ne lehteen sopivaksi sekä suorittaa muita jatkotoimenpiteitä. Ylläpitöhenkilöt voivat myös vastata artikkelien DTD:eistä. Näitä kaikkia toimintoja ei toteuta tuotteemme editorilla vaan niitä voidaan käyttää esimerkiksi tietokannasta käsin.

JouTo-järjestelmän ylläpito olisi lähinnä tietokannan ylläpitoa. Järjestelmän ylläpitäjän käyttötapauksia tukevia ominaisuuksia ei kuitenkaan toteuteta.

Niitä olisivat ainakin:

- käyttöprofiilien lisäys, muokkaus ja poisto
- vanhentuneiden artikkelien tuhoaminen
- dokumenttityyppimäärittelyjen lisäys, muokkaus ja poisto

2.3 Toteutusympäristö

Demoversio ohjelmistosta toimi Helsingin yliopiston tietokonejärjestelmässä sekä Linux- että Windows-ympäristöissä, jossa ohjelmiston kehittäminen on toteutettu. Demoversio käyttää tietokantaa, joka on Oracle 8-pohjainen ja sijaitsee Helsingin Yliopiston atk-keskuksen *kontti*-tietokantapalvelimella.

Ohjelmistoa käytetään WWW-selaimen kautta. Ohjelmiston asiakkaan puolen käyttöliittymä, editori, toteutettiin WWW-selaimessa toimivana graafisena Java applettina. Appletissa käytetään Javan version 1.2 ominaisuuksia ja valmiita komponentteja, kuten Swing-kirjastoa. Lisäksi käytetään JouTo-projektin ulkopuolella toteutettuja Java-kirjastoja, Xerces XML-parseria, Mark Wutkan tekemää DTD-parseria ja aelfred-parseria. Näiden lisäksi ohjelmistossa luodaan omia tietorakenteita sekä XML-puun käsittelyä että varsinaisen artikkelin muokkausta varten.

Tietokanta -operaatioista ja tietokannan käsittelystä huolehtivat JouTo-projektissa luodut Java Servlet-komponentit. Nämä komponentit käyttävät JDBC-rajapintaa tietokannan käsittelyssä. JDBC-tietokantarajapinta kuuluu osana Java 1.2 -standardiin.

2.4 Toimintaympäristö

JouTo-projektin editori toimii siinä ympäsitössä, jossa tuotteen kehittäminen on toteutettu.

Jos demoversiosta kehitetään varsinainen tuote, tällöin toimintaympäristöksi on oletettu toimituksen sisäinen verkko. Uutistoimittajat kirjoittavat artikkeleitaan tietokoneen äärellä usein toimittajan työpaikalla, toimituksessa. Toimittajilla on käytössään tietokoneita, jotka ovat yhteydessä verkon sisällä olevaan palvelimeen. Artikkelit kirjoitetaan ja muokataan tietokoneella ja tallennetetaan sekä ladataan palvelimella olevaan tietokantaan. Koska kyseessä on tomituksen sisäinen verkko, voimme olettaa yhteyksien olevan suojattuja.

2.5 Asennustiedot

2.6 Järjestelmän liittyvät tiedostot ja lähdekoodi

3 Suunnitteluun toteutuksen aikana tehdyt muutokset ja rajoitukset

Järjestelmän suunnittelu on kuvattu suunnitteludokumentissa. Tässä luvussa esitellään toteutuksen aikana suunnitteluun tehdyt muutokset pääpiirteissään.

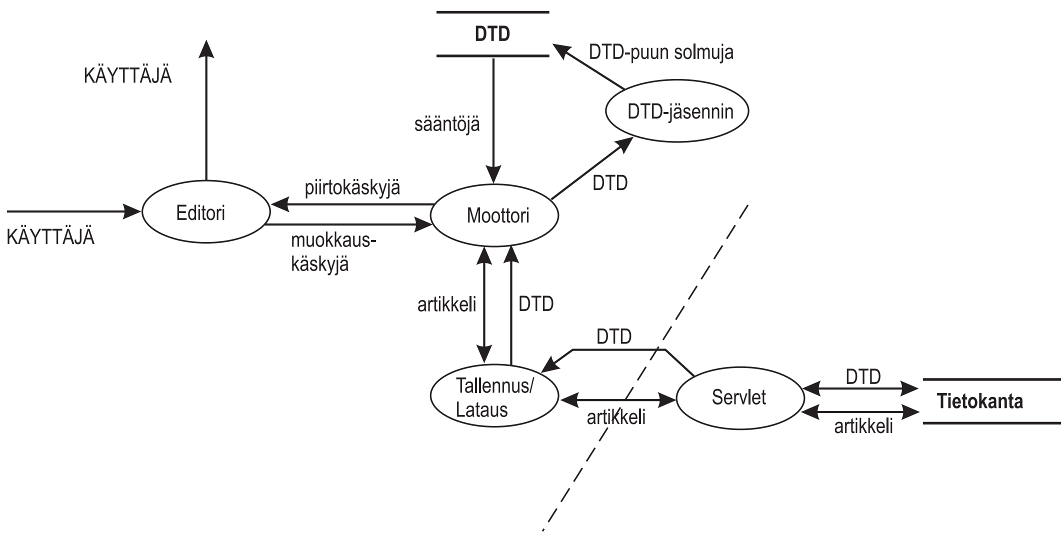
3.1 Toimintaympäristö

Järjestelmä suunniteltiin alunperin toimimaan selaimessa toimivana applettina. Toteutuksen aikana huomattiin, että Tietojenkäsittelytieteen laitoksen ATK-ympäristön Java Runtime Environment oli väärässä versiossa, eikä tarvittavaa versiota saatu asentumaan. Näin ollen ohjelmaa ei voitu ajaa laitoksen koneiden selaimissa. Ohjelma toteutettiin toimimaan siten myös omana sovelluksenaan.

Tietokanta saatiin toimimaan suunniteludokumentissa selvitetyllä tavalla Tietojenkäsittelytieteen tietokantapalvelmilla (db.cs.helsinki.fi).

3.2 Yleisarkkitehtuuri

Yleisarkkitehtuurissa ei ole tehty muutoksia. Arkkitehtuuri kuvataan kuvassa 1. Yleinen arkkitehtuurikuvaus on esitelty ja selitetty suunnitteludokumentissa.



Kuva 1: Ohjelmiston arkkitehtuurikuvaus

3.3 Luokkarakenne

Järjestelmän luokkarakenteen muutokset on kuvattu kappaleessa 4 *Järjestelmän luokkarakenne*.

3.4 DTD:hen tehdyt muutokset

Suunnitteluvaiheessa päätettiin toteuttaa JouTo-projektia varten oma DTD, joka on rajoitettu versio NITF v.2.5 DTD:stä. Sovelluksessa käytettävä DTD luotiin JouTo-järjestelmän toimintoja vastaaviksi. Toteutettu DTD on kuvattu liitteessä (Liite A).

Järjestelmän ideana oli luoda mahdollisuus käyttää myös muita DTD:eitä. Toteuksen aikana luotiin siten toteutusta DTD:stä vielä rajoitetumpi DTD kolumnien (esim. pääkirjoituksia) varten. Myös tämä DTD löytyy liitteistä (Liite B).

4 Järjestelmän luokkarakenne

Tässä luvussa esitellään toteutetun järjestelmän luokat. Ensin luokkarakenteesta luodaan yleiskuva luokkakaavon avulla. Tämän jälkeen seuraa tarkemmat luokkakuvaukset, joissa esitellään jokaisen luokan osalta sen tarkoitus, attribuutit ja metodit.

4.1 Luokkakaavio

4.2 Luokkien kuvaukset

5 Jatkokehitysajatuksia

Toteutettu järjestelmä on toimiva yhdistyksen WWW-järjestelmän perusrunko. Tässä luvussa esitellään muutamia ajatuksia järjestelmän jatkokehittelylle.

5.1 Turvallisuusnäkökohdat

5.2 Ylläpidettävyys

5.3 Järjestelmän toiminnot

A Liite: Toteutettu DTD

```
<!ENTITY % enriched-text '
        #PCDATA
        | location
        | object.title
        | org
        | person
        | a
        | br
        | em
        | q
        '>
<!ENTITY % block.content "p | hl2 | media | ol | ul | pre ">

<!ELEMENT nitf (head, body)>
<!ATTLIST nitf
        version      CDATA    #FIXED    "-//IPTC-NAA//DTD NITF-XML 2.1//EN"
        change.date  CDATA    #FIXED    "4 July 2000"
        change.time  CDATA    #FIXED    "1900"
        >
<!ELEMENT head (title, tobject, docdata)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT tobject (tobject.subject)>
<!ATTLIST tobject
        tobject.type            CDATA   "news"
        >
        <!ELEMENT tobject.subject  EMPTY>
        <!ATTLIST tobject.subject
                tobject.subject.ipr            CDATA          "IPTC"
                tobject.subject.refnum         NMTOKEN        #REQUIRED
                >
<!ELEMENT docdata (
        evloc,
        doc-id,
        date.issue,
        doc.copyright?,
        key-list
        )>
        <!ELEMENT evloc EMPTY>
        <!ATTLIST evloc
                city            CDATA   #IMPLIED
                >
        <!ELEMENT doc-id EMPTY>
        <!ATTLIST doc-id
                id-string       CDATA   #IMPLIED
```

```
                >
        <!ELEMENT date.issue EMPTY>
        <!ATTLIST date.issue
                norm            CDATA   #IMPLIED
                >
        <!ELEMENT doc.copyright EMPTY>
        <!ATTLIST doc.copyright
                year            NMTOKEN         #IMPLIED
                holder          CDATA           #IMPLIED
                >
        <!ELEMENT key-list (keyword)*>
                <!ELEMENT keyword EMPTY>
                <!ATTLIST keyword
                        key             CDATA   #REQUIRED
                        >
<!ELEMENT body (body.head, body.content)>
<!ELEMENT body.head (
        hedline,
        byline,
        distributor?,
        abstract
        )>
<!ELEMENT hedline (hl1, hl2*)>
        <!ELEMENT hl1 (%enriched-text;)*>
        <!ELEMENT hl2 (%enriched-text;)*>
<!ELEMENT byline (#PCDATA)>
<!ELEMENT distributor (#PCDATA)>
<!ELEMENT abstract (p)>
<!ELEMENT body.content (%block.content;)*>
<!ELEMENT p (%enriched-text;)*>
<!ELEMENT media (media-reference, media-caption?)>
<!ATTLIST media
        media-type      CDATA #FIXED "image"
        >
        <!ELEMENT media-reference (#PCDATA)>
        <!ATTLIST media-reference
                data-location           CDATA           #REQUIRED
 mime-type  CDATA   #REQUIRED
                >
        <!ELEMENT media-caption (%enriched-text;)*>
<!ELEMENT ol (li)+>
<!ELEMENT ul (li)+>
        <!ELEMENT li (%enriched-text;)*>
<!ELEMENT pre (#PCDATA)>
<!ELEMENT location (#PCDATA)>
<!ELEMENT object.title (#PCDATA)>
```

```
<!ATTLIST object.title
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT org (#PCDATA)>
<!ATTLIST org
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT person (#PCDATA)>
<!ATTLIST person
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT a (%enriched-text;)*>
<!ATTLIST a
        href            CDATA           #IMPLIED
        name            CDATA           #IMPLIED
        title           CDATA           #IMPLIED
        >
<!ELEMENT br EMPTY>
<!ELEMENT em (%enriched-text;)*>
<!ELEMENT q (%enriched-text;)*>
<!ATTLIST q
        quote-source    CDATA   #IMPLIED
        >
```

B Liite: Kolumni DTD

```
<!ENTITY % enriched-text '
        #PCDATA
        | location
        | object.title
        | org
        | person
        | a
        | br
        | em
        | q
        '>
<!ENTITY % block.content "p | hl2 ">

<!ELEMENT nitf (head, body)>
<!ATTLIST nitf
        version      CDATA    #FIXED    "-//IPTC-NAA//DTD NITF-XML 2.1//EN"
        change.date  CDATA    #FIXED    "4 July 2000"
        change.time  CDATA    #FIXED    "1900"
        >
<!ELEMENT head (title, docdata)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT docdata (
        doc-id,
        date.issue,
        doc.copyright?,
        key-list
        )>
        <!ELEMENT doc-id EMPTY>
        <!ATTLIST doc-id
                id-string       CDATA   #IMPLIED
                >
        <!ELEMENT date.issue EMPTY>
        <!ATTLIST date.issue
                norm            CDATA   #IMPLIED
                >
        <!ELEMENT doc.copyright EMPTY>
        <!ATTLIST doc.copyright
                year            NMTOKEN         #IMPLIED
                holder          CDATA           #IMPLIED
                >
        <!ELEMENT key-list (keyword)*>
                <!ELEMENT keyword EMPTY>
                <!ATTLIST keyword
                        key             CDATA   #REQUIRED
```

```
                        >
<!ELEMENT body (body.head, body.content*)>
<!ELEMENT body.head (
        hedline,
        byline,
        distributor?,
        abstract
        )>
<!ELEMENT hedline (hl1, hl2*)>
        <!ELEMENT hl1 (%enriched-text;)*>
        <!ELEMENT hl2 (%enriched-text;)*>
<!ELEMENT byline (#PCDATA)>
<!ELEMENT distributor (#PCDATA)>
<!ELEMENT abstract (p)>
<!ELEMENT body.content (%block.content;)*>
<!ELEMENT p (%enriched-text;)*>
<!ELEMENT media (media-reference, media-caption?)>
<!ATTLIST media
        media-type      CDATA #FIXED "image"
        >
        <!ELEMENT media-reference (#PCDATA)>
        <!ATTLIST media-reference
                data-location           CDATA           #REQUIRED
 mime-type  CDATA   #REQUIRED
                >
        <!ELEMENT media-caption (%enriched-text;)*>
<!ELEMENT location (#PCDATA)>
<!ELEMENT object.title (#PCDATA)>
<!ATTLIST object.title
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT org (#PCDATA)>
<!ATTLIST org
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT person (#PCDATA)>
<!ATTLIST person
idsrc  CDATA  #REQUIRED
value  CDATA  #REQUIRED
>
<!ELEMENT a (%enriched-text;)*>
<!ATTLIST a
        href            CDATA           #IMPLIED
        name            CDATA           #IMPLIED
```

```
        title           CDATA           #IMPLIED
        >
<!ELEMENT br EMPTY>
<!ELEMENT em (%enriched-text;)*>
<!ELEMENT q (%enriched-text;)*>
<!ATTLIST q
        quote-source    CDATA   #IMPLIED
        >
```