Määrittelydokumentti

WWW-museo

Helsinki 20. helmikuuta 2002

Versio 1.1

HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Sisältö

1 Johdanto

Projektissa on tarkoitus tuottaa selainpohjainen dynaaminen WWW-sivusto, eli virtuualimuseon näyttely. Näyttelyssä vierailija voi käydä läpi tietokannassa olevia kuvia. Näyttelyn on tarkoitus olla älykäs, eli ohjelman pitäisi osata päätellä vierailijan kiinnostuksen kohteet ja ehdottaa sopivia kuvia. Näyttelyn teemana on promootio.

Tässä dokumentissa käsitellään toteutettavaa ohjelmistoa hyvin yleisellä tasolla. Projektissa käytetään monia uusia välineitä ja tekniikoita, joiden toiminta ja ominaisuudet eivät ole ennestään tuttuja. Niinpä yksityiskohtaiset suunnitelmat ohjelmiston eri osista esitetään vasta suunnitteludokumentissa, jolloin ryhmällä on ollut riittävästi aikaa tutustua projektissa käytettäviin välineisiin ja tekniikoihin. Tässä dokumentissa esitetään vain lähtökohdat suunnittelulle.

2 Termit, määritelmät ja lyhenteet

Termi:	Määrittely:
DBXML	XML-tietokanta. Tarkempaa tietoa osoitteesta www.dbxml.org/.
CSS	Cascading Style Sheets. Merkkauskieli, jolla voidaan määritellä esim. HTML dokumenttien ulkoasu. Tarkempaa tietoa osoitteessa www.w3.org/Style/CSS/.
CVS	Concurrent Versions System. Versionhallintaohjelma.
Fasetti	Hakukäsite. Muodostuu saman aihealueen sisällä samaa aspektia kuvaavista asioista tai niitä tarkoittavista käsitteistä. Niiden käsitteiden joukko, joita voidaan pitää haussa vaihtoehtoisina. Fasetti on siis käsiteryväs.
HTML	Hypertext Markup Language. WWW-sivujen kuvaamiseen käytettävä kieli. Tarkempi kuvaus osoitteessa www.w3.org/TR/html4/.
JavaBeans	Teknologia, jolla voidaan luoda ohjelmakomponentteja Java-ympäristössä. Katso lisää osoitteesta http://java.sun.com/products/javabeans/docs/.
JavaScript	Selaimen tulkitsema kieli, jolla voidaan lisätä HTML-sivuihin toimintoja, joita pelkkä HTML ei mahdollista.
JDBC	Java-kielen rajapinta, jonka avulla voidaan käsitellä tietokantoja.
Jena	Javalla toteutettu työkalu, jolla voidaan käsitellä RDF-malleja. Tarkempi kuvaus osoitteessa www.hpl.hp.com/semweb/jena-top.html.
JRE	Java Runtime Environment. Sunin kehittämä ohjelmointiympäristö. Tarkempaa tietoa osoitteessa http://java.sun.com/.

JSP	JavaServer Pages. Käyttää XML:n kaltaisia tageja kapseloimaan logiikan, joka luo web-sivun. Tarkempaa tietoa osoitteessa http://java.sun.com/products/jsp/.
Ontologia	Tiedon riippuvuussuhteiden esittäminen puumaisena rakenteena. Puun juurena on esim. talo, jonka lapsia ovat ylä- ja alakerta ja lapselapsia ullakko ja keittiö.
RDF	Resource Description Framework. Kieli, jolla voidaan yhdistää erilaista XML-muodossa olevaa tietoa toisiinsa ontologioiden avulla. Tarkempaa tietoa osoitteessa www.w3.org/RDF/.
SQL	Structured Query Language. Kieli, jolla voidaan muokata ja käyttää tietokantaa.
Vanha	Helsingin keskustassa sijaitseva Vanha Ylioppilastalo. Talossa on mm. ravintola ja siellä järjestetään juhaltilaisuuksia. Katso lisää ositteesta www.vanha.fi.
Xerces	XML-parseri XML-dokumenttien käsittelyyn.
XHTML	Extensible Hypertext Markup Language. XML-yhteensopiva HTML-standardia mukaileva WWW-sivujen kuvauskieli.
XML	Extensible markup language. Metakieli rakenteisten dokumenttien esittämiseen. Tarkempaa tietoa osoitteessa www.w3.org/TR/2000/REC-xml-20001006.

3 Sidos- ja käyttäjäryhmät

Järjestelmällä on viisi sidosryhmää (Kuva 1). Seuraavassa sidosryhmien kuvaukset:

Vierailijat– Vierailijoilla tarkoitetaan ketä tahansa museon www-sivuille sattuvaa, promootiosta kiinnostunutta henkilöä, jota ensisijaisesti kiinnostaa mielekkäällä tavalla esitetty kuvallinen esitys promootiosta.

Tutkijat– Tutkija tarvitsee täsmätietoa tarkkaan määritellystä promootioaineistoon liittyvästä tekijästä. Tutkija saattaa olla kiinnostunut esimerkiksi vain tietyn aikakauden promootioista, tietyistä henkilöistä (rehtori, promovoitava jne.) tai jostain tietystä rakennuksesta.

Museon työntekijät– Museon työntekijät keräävät aineiston näyttelyä varten ja vastaavat sen siirrosta digitaaliseen muotoon sekä ylläpidosta. Näyttely koostuu pääosin kuvista ja niihin liittyvästä metatiedosta. Kuvien lisäksi näyttelyyn kuuluu promootion sanallista kuvausta, esimerkiksi yleiskuvaus, tapahtumien ja eri henkilöiden tehtävien kuvaukset. Museon työntekijät ovat siis vastuussa järjestelmän tietosisällöstä.

Tässä projektissa ei keskitytä tietosisällön ylläpitämiseen, vaan lähtökohtana on jo olemassaoleva kuva- ja metatietotietokanta, josta valitaan promootioon liittyvät osat. Uusien näyttelyiden luominen ja puutteellisten tai virheellisten tietojen korjaaminen toteutetaan mahdollisissa jatkoprojekteissa.

Tietojenkäsittelytieteen tutkijat– Tietojenkäsittelytieteen tutkijoita kiinnostaa itse ohjelman toiminta. Olennaista on mahdollisuus tutkia uudella tekniikalla toteutetun järjestelmän toimintaa ja toimivuutta. Tutkija haluaa tietää, miten järjestelmän mielenkiintoisten kuvien päättely etenee ja kuinka hyvin tässä onnistutaan. Myös erilaiset tilastotiedot järjestelmän käytöstä ovat mahdollisesti tutkijan kiinnostuksen kohteita.

Ylläpitäjät– Ylläpitäjillä tarkoitetaan tässä yhteydessä lähinnä jatkokehittäjiä, sillä toteutettava järjestelmä ei sellaisenaan ole tarkoitettu ylläpitoa vaativaan tuotantokäyttöön. Olennaisin puute todellisen käytön kannalta on järjestelmän tietosisällön ylläpito-ominaisuuksien puute.

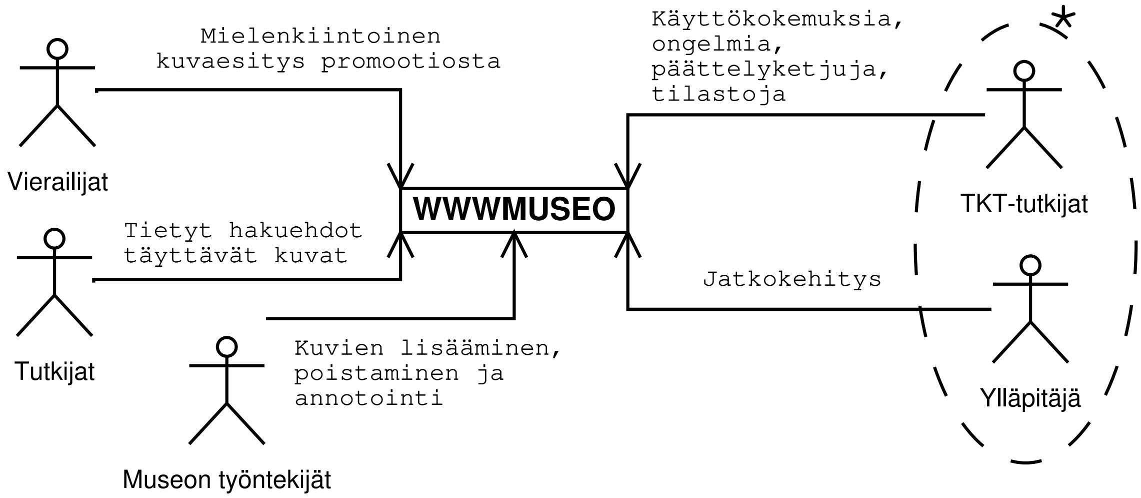
4 Ohjelmistolle asetetut vaatimukset

4.1 Järjestelmän tarjoamat palvelut

Järjestelmä tarjoaa käyttäjille seuraavat palvelut:

4.1.1 HiBrowse

HiBrowse on käyttöliittymän osa, jonka avulla käyttäjä voi tarkentaa hakua kuvatietokannasta. Ensin käyttäjä hakee esimerkiksi kaikki kuvat Vanhasta. Tämän



Kuva 1: WWW-museon sidosryhmät. *) Ohjelman jatkokehittäjät ovat todennäköisesti myös TKT:n tutkijoita tai ainakin jatkokehitys tehdään heidän tulostensa pohjalta.

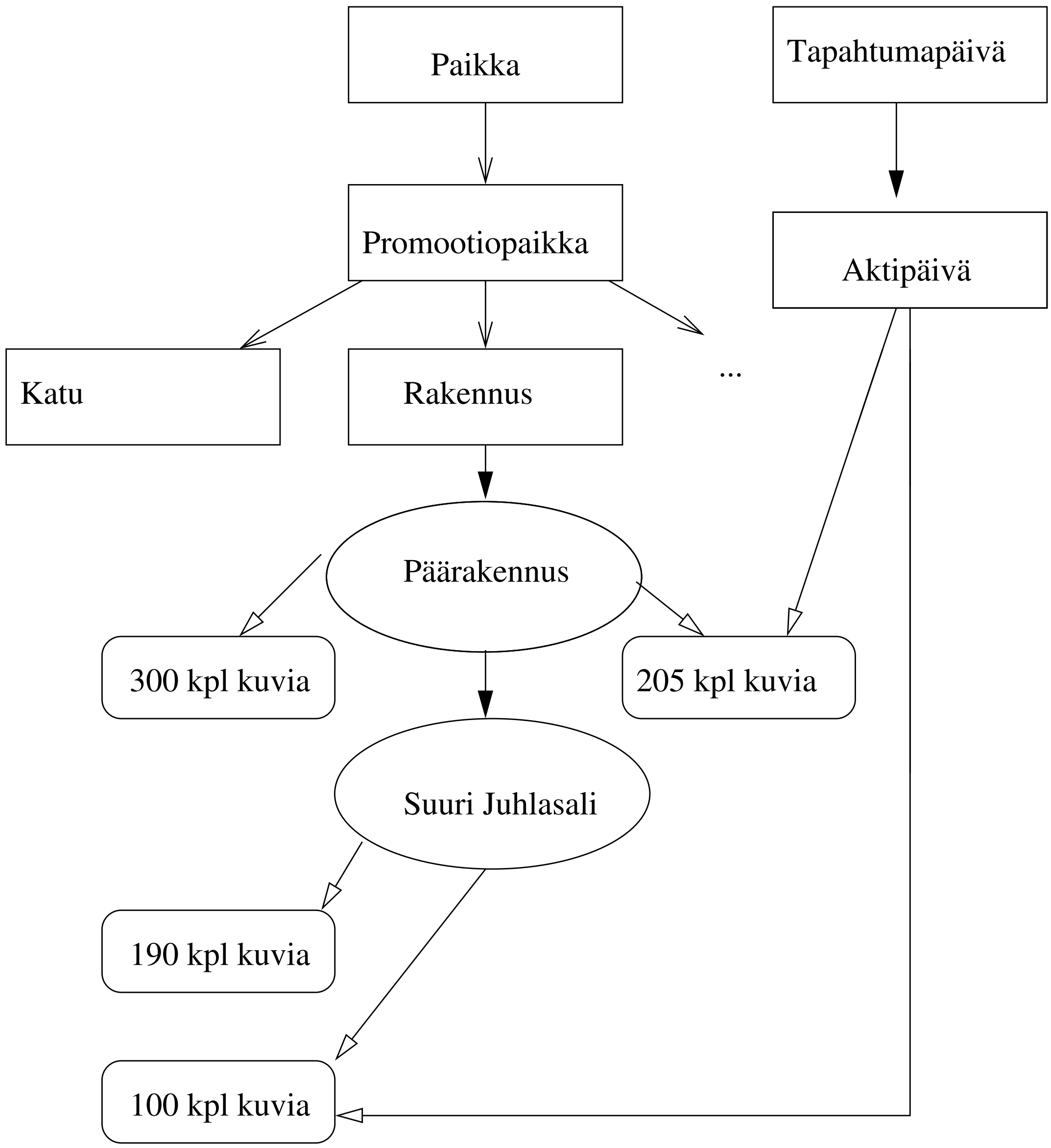
jälkeen hän tarkentaa hakua Vanhan yläkerrasta otettuihin kuviin. Tämän jälkeen hakua voidaan vielä tarkentaa tiettyyn yläkerran huoneeeseen. Haku tehdään fasettien avulla. Fasettien ja ontologioiden toimintaa havainnollistetaan kuvassa 2. Faseteilla tehdyt valinnat rajaavat tietokannasta löytyvien kuvien joukkoa. Mitä enemmän valintoja tehdään ja mitä syvemmälle ontologiassa mennään, sitä vähemmän halutunlaisia kuvia löytyy.

4.1.2 Mielenkiintoiset linkit

Tämä palvelu toteuttaa järjestelmän tekoälyn. Kun käyttäjä katsoo jotakin kuvaa, kuvan viereen tulee linkkilista muista kuvista, jotka saattavat myös kiinnostaa käyttäjää. Näissä muissa kuvissa saattaa olla sama henkilö, sama paikka tai vaikkapa sama aika kuin käyttäjän valitsemassa kuvassa.

4.1.3 Tapahtumapohjainen suunnistaminen

Palvelu, jonka avulla käyttäjä voi etsiä kuvia promootiotapahtuman eri vaiheista.



Kuva 2: Ontologiat ja fasetit

4.1.4 Peruuta–toiminto

Tämän palvelun avulla käyttäjä pääsee valinnoissaan yhden askeleen taaksepäin. Tästä on apua jos käyttäjä painaa vahingossa väärää linkkiä tms.

4.1.5 Vapaamuotoinen haku

Tämän palvelun avulla käyttäjä voi hakea kuvatietokannasta kuvia haluamallaan hakusanalla. Esimerkiksi hakusanalla "miekka"näytetään kaikki kuvat, joissa on miekka.

4.2 Lisäominaisuuksiin varautuminen

—Käyttäjäprofiili

Ohjelmistoon tulee olla mahdollista lisätä myöhemmin käyttäjäprofiilit ja niiden avulla tehtävä kuvien ja älykkäiden linkkien rajaus. Kuvien ja linkkien rajaamis/arvottamisfunktio suunnitellaan modulaariseksi ja laajennettavaksi.

—Laitteistoprofiili ja monikielisyys

Laitteistoprofiilien ja monikielisyyden tueksi käyttöliittymä ja muu koodi erotetaan toisistaan mahdollisimman hyvin. JSP -sivuissa käytetään tagikirjastoja.

—Istuntohistoria

Istuntohistorian tallentamisen mahdollistamiseksi ohjelmisto yritetään suunnitella niin, että sen tila on helppo tallentaa kullakin hetkellä ja että kulloinkin voimassa oleva tila on aina ohjelman tiedossa (varaudutaan back-nappulan painamiseen ja muihin komplikaatioihin).

4.3 Järjestelmän tekninen toteutus

Järjestelmä toteutetaan Java Server Pages –järjestelmällä. Java-koodit erotetaan sivuista Java-papujen (engl. JavaBeans) ja JSP-merkkausten, eli tagien, avulla.

Promootiokuvista generoidaan XML-tietovarasto. Kuvien tietoja täydennetään mahdollisuuksien mukaan.

XML-tietovaraston ja RDF-ontologioiden perusteella luodaan tiedonhakujärjestelmä. Jäjestelmä toteutetaan joko keskusmuistipohjaisena tai tietokantapohjaisena, jolloin tietokantana käytetään järjestelmää nimeltä DBXML.

4.4 Käyttäjäryhmät

—Vierailijat

Käyttöliittymän on oltava satunnaisen museonkävijän ymmärrettävissä. Käyttöliittymä suunnitellaan mahdollisimman helppokäyttöiseksi. Monimutkaisemmat toiminnot piilotetaan perusnäkymästä.

—Tutkijat

Käyttöliittymästä tulee myös löytyä oikoteitä kokeneelle ohjelmiston käyttäjälle. Ohjelma suunnitellaan niin, että myös sen monimutkaisemmat valintamekanismit voidaan antaa käytettäviksi.

—Yleistettävyys

Ohjelmisto suunnitellaan siten, että sitä on helppo mukauttaa uuteen ympäristöön, muihin näyttelyihin ja kokonaan muihin tarkoituksiin. Ohjelmiston komponenttien tulee olla mahdollisimman modulaarisia ja yleispäteviä.

5 Kuvaus käyttöliittymästä

Käyttöliittymä on keskeinen osa projektin vaatimuksia, joten se käsitellään yksityiskohtaisesti vasta suunnitteludokumentissa. Seuraavaksi mainitaan muutamia käyttöliittymän perusvaatimuksia.

5.1 Käyttäjäryhmät

Käytettävyyttä suunniteltaessa otetaan huomioon seuraavat käyttäjäryhmät. Ryhmät ovat tärkeysjärjestyksessä.

1. Satunnainen museossa kävijä/ webin selaaja

On henkilö, jolla ei välttämättä ole minkäänlaista tietoa promootiosta ja tarjolla olevista kuvista.

2. Tutkijat

On henkilö, joka tuntee kuvakokoelmaa ja on usein tietoinen siitä, mitä etsii.

5.2 Toteutettavat osat

1. Monifasettihaku

Mahdollistaa kuvien filtteröinnin usean eri fasetin avulla samanaikaisesti.

2. Kuvalista

Lista kuvista, jotka sopivat hakuehtoihin.

3. Kuvasivu

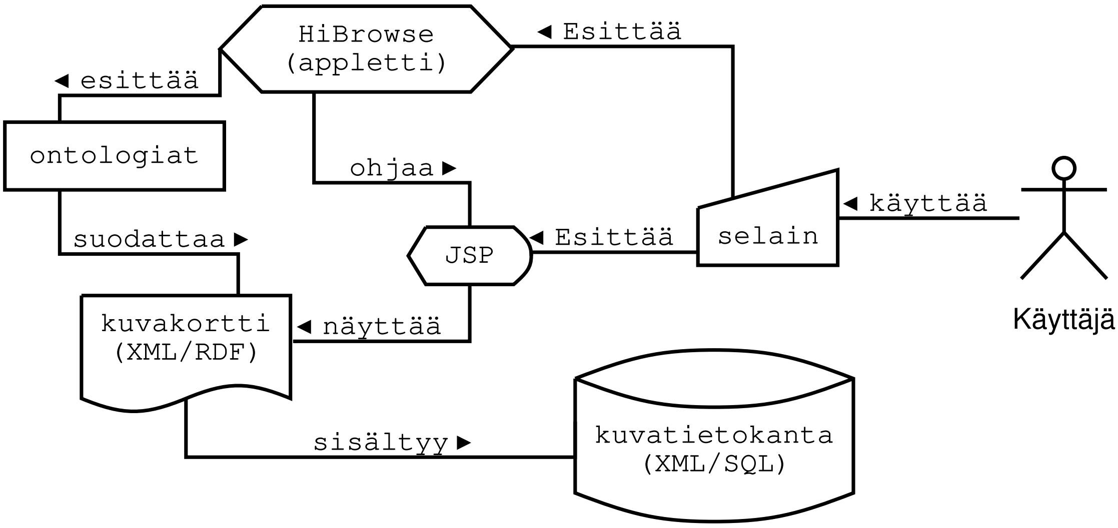
Kuvasivu sisältää itse kuvan, tiedot kuvista ja linkkejä muihin käyttäjää kiinnostaviin kuviin.

5.3 Toteutustekniikat

JavaApplet, JSP, JavaScript ja HTML.

5.4 Erityisvaatimukset

Järjestelmän on toimittava yleisimmissä selaimissa, joissa on Java 1.3 plug-in. Käyttäjältä ei voida vaatia minkäänlaisia erikoistaitoja. Selaimen peruskäyttötaidot riittävät.



Kuva 3: Ulkoiset liittymät

6 Järjestelmän ulkoiset liittymät

6.1 Käyttöliittymä

Kuvattiin luvussa 4. Tarkempi määrittely vasta suunnitteludokumentissa.

6.2 Laitteistoliittymät

Ohjelmisto on laitteistoriippumanton. Toteutettavat ohjelmistot toimivat javan ajonaikaisen ympäristön versiossa 1.3.1 (Java Runtime Environment, JRE). Ohjelmassa on kaksi osaa, yksi käyttäjän selaimessa toimiva appletti joka vaatii toimiakseen JREn mukana tulevan java-pluginin, ja palvelimella toimiva kuvia näyttävä kokonaisuus joka vaatii että palvelimelle on asennettu JRE.

6.3 Ohjelmistoliittymät

Ohjelmistoliittymät on esitetty kuvassa 3.

Käyttöliittumänä toimivan HiBrowse-mallin (luku 5.1) mukainen ontologioiden selailujärjestelmä on selaimessa toimivan java-pluginin alaisuudessa toimiva sovelma (applet). Erityisesti siitä käytetään swing-käyttöliittymäkirjastoa. Muita selaimeen kohdistuvia vaatimuksia on tuki HTML-kuvauskielen versiolle 4.01 tai XHTML-kuvauskielen versiolle 1.0. Myös tuki CSS-tyylikuvauskielen versiolle 1 vaaditaan. Mahdollisten XML-tiedostojen käsittelyyn vaaditaan Apache-projektin (www.apache.org) kehittämää, vapaata ja ilmaista, Xerces-ohjelmakirjastoa. RDF-dokumenttien käsittelyyn vaaditaan HP-Labsin kehittämää, vapaata ja ilmaista, Jena-ohjelmakirjastoa. Palvelinpuolella ajettavat ohjelmat ovat ns. java-papuja (beans) ja ne ajetaan JSP-ympäristössä. Varsinainen JSP-ympäristön tarjoava palvelinohjelmisto koostuu Tomcat JSP-palvelimesta joka toimii Apache-www-palvelinohjelmisto alaisuudessa. Molemmat ovat Apache-projektin kehittämiä, vapaita ja ilmaisia ohjelmia. Tietokantajärjestelmänä käytetään yksinkertaista tiedostoa tai JDBC:llä ohjattua SQL-pohjaista tietokantajärjestelmää.

6.4 Tietoliikenneliittymät

Tietoliikenne selaimessa toimivan sovelman ja palvelinpuolella toimivien JSP-papujen välillä hoidetaan http-protokollan ja URLien välityksellä. Erillistä tietoliikenneratkaisua ei näillä näkymin tarvitse toteuttaa. Jos oman TCP-pohjaisen tietoliikenneliittymän toteuttaminen tulee ajankohtaiseksi, kyseeseen tulee lähinnä hyvin yksinkertainen asiakas/palvelinratkaisu säikeillä ja TCP-paketeilla.

6.5 Alustustiedostot

Ohjelmiston pitäisi pystyä tekemään kaikki vaadittavat asetukset valmiiden, palvelimella olevien ontologioiden perusteella. Mahdolliset käyttäjäprofiilit (luku 3.1 - lisäominaisuuksiin varautuminen) voidaan toteuttaa joko asiakaspuolella keksien (cookies) avulla tai palvelinpuolella.

6.6 Laitteistorajoitukset

Ohjelmiston tulisi toimia kohtuullisen hyvin nykyaikaisessa selaimessa, tietokoneella jonka työpöydän resoluutio on vähintään 800*600, mieluiten 1024*768. Selaimessa tulee olla käytettävissään riittävä määrä värejä kuvien tyydyttävään näyttämiseen (väh. 16-bittiset värit). Suorituskyyltään ohjelmiston tulee pystyä palvelinpuolella esittämään n. 30000 kuvan kokoelmaa nykymittapuun mukaan kohtalaisen nopeissa (686-tasoisissa tai vastaavissa) työasemissa ja palvelimilla.

7 Käyttötapaukset

7.1 Tutkijat

Tutkija T tekee elämänkertaa professorista P, haluten löytää promootiokuvia professorista. Mikäli tekstihaku on toteutettu, syöttää hän haettavaksi professorin nimen. Mikäli vain fasettihaku on mahdollinen, valitsee T hierarkiasta Henkilö/Promootiohenkilö/ . Nyt hän huomaa että ontologiahierarkiassa on valittavana kaksi vaihtoehtoa: Maisteripromovendi ja Tohtoripromovendi. Koska T tekee elämänkertaa päättää hän etsiä kummastakin oksasta kuvia P:stä, jonka promovoinneista hän on kiinnostunut. T ei siis välttämättä ole kiinnostunut muiden ihmisten promootioista, joissa P on ollut esim. seuralaisena tai sukulaisena.

7.2 Vierailijat

Vierailija V selailee satunnaisesti promootion vaihekarttaa. Promootiopäivällisen kohdalla hän kiinnostuu siitä missä eri ravintoloissa ko. päivällisiä on pidetty ja ennen kaikkea, että niistä olisi kuvia. Kuvia etsiäkseen hän valitsee tapahtumafasetista Päivällisen. Tämän jälkeen kuvat on rajattu päivälliseen liittyviin kuviin.

7.3 Museon työntekijät

Työntekijä M haluaa lisätä vastikään promovoidun tohtorin kuvia tietokantaan. Hän valitsee ohjelmistosta toiminnon, jolla voi syöttää uusia kuvia tietoineen kantaan. Jokaisesta kuvasta hän kirjoittaa seuraavat tiedot: paikka, päivämäärä, kellonaika, kuvassa esiintyvät henkilöt, tilaisuus (päivällinen, miekanhiojaiset...).

7.4 Tietojenkäsitteelytieteen tutkijat

Tästä ryhmästä on vaikea keksiä edustavaa käyttötapausta, sillä tutkimuskohteita on useita. Tutkija K haluaa tutkia ontologiaa, tällöin hän valitsee ontologiahierarkian yleiskuvauksen (puumuodossa). Mikäli tällaista ei ole saatavilla, selvittää hän sen fasettien avulla ontologiahierarkiassa navigoimalla.

7.5 Ylläpitäjät

Tämä sidosryhmä liittyy tietyllä tapaa edelliseen ryhmään: kummatkin ovat kiinnostuneita ohjelman toiminnasta ja logiikasta. Erona kuitenkin on että Ylläpitäjien tarkoitus on muuttaa järjestelmän toimintaa esim. laajentamalla sitä. Yksi laajennus voisi olla videoiden lisääminen tietokantaan. Tämä edellyttää uudenlaisen "XML-kortin", joka liittyy jokaiseen mediaobjektiin, luomista. Tähän korttiin tulisivat seuraavat tiedot: videon kesto, väri/mustavalko, ääni jne. Myös liittymän rakentaminen videontoisto-ohjelman on tehtävä.