Asiakas: Kjell Lemström　　　　　　Ohjaaja: Anita Kasari

pics98

Vaatimusmäärittely

Samuli Kaski
Timo Koskinen
Jani Leinonen
Mika Pennanen

Helsinki 2 elokuuta 1998

Ohjelmistotuotantoprojekti, kesä 1998
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Sisältö

1 Johdanto (TK)

PICSearch on ensisijaisesti asiantuntijakäyttöön suunnattu kuvan sisältöön perustuvien hakualgoritmien kehitysohjelmisto, joka tarjoaa valmiin testausalustan algoritmeille.

Järjestelmän perustoteutus, PICSearch v1 (ks. [1]), joka mahdollistaa kuvien haun tietokannasta esimerkkikuvien perusteella on toteutettu ohjelmistotyöprojektina Helsingin teknillisessä korkeakoulussa. Helsingin yliopiston NetPICSearch-ohjelmistotuotantoprojektissa (ks. [2]) ohjelmistoon lisättiin verkkotuki ja korjattiin joitakin edelliseen versioon jääneitä virheitä. Helsingin yliopiston `pics98`-ohjelmistotuotantoprojektin päätarkoitus on korjata järjestelmään jääneitä virheitä, parantaa ohjelman antamaa palautetta ja uudistaa käyttöliittymä.

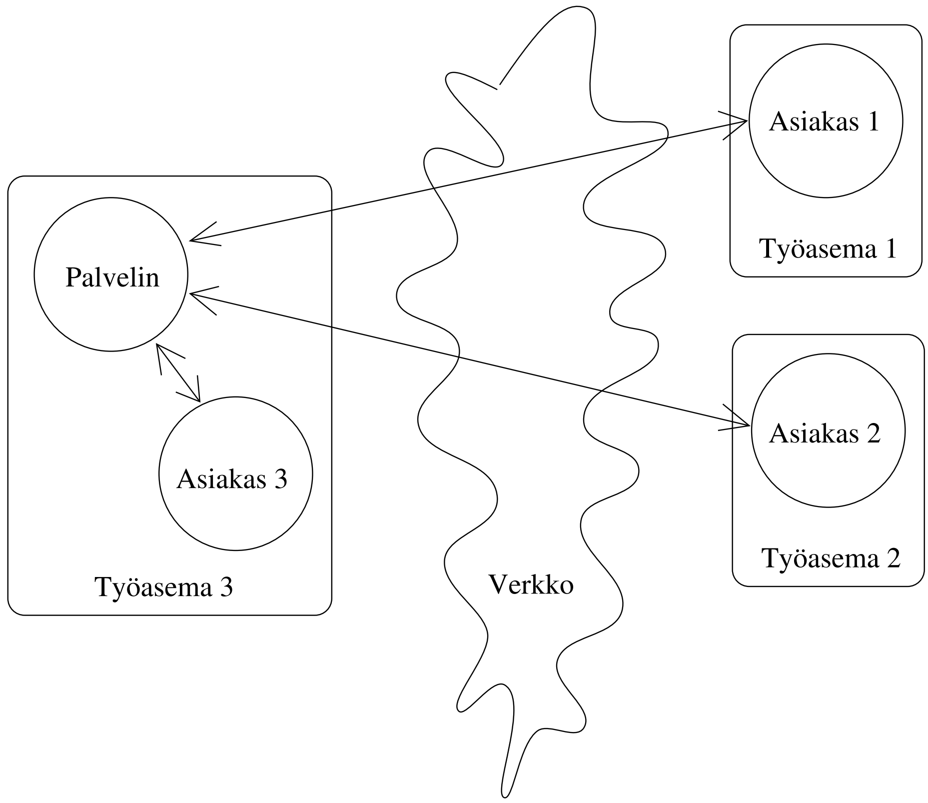
Tässä dokumentissa määritellään ohjelmistolle asetettavat vaatimukset. Lukijan oletetaan tuntevan PICSearchin toimintaperiaate. Luvussa 2 kuvataan järjestelmän toiminta, ympäristö ja ohjelmistoon tehtävät muutokset. Luvussa 3 määritellään järjestelmän tietosisältö ja luvussa 4 tiedonvälitys. Luvussa 5 kuvataan käyttöliittymään tehtävät parannukset ja uudet ominaisuudet. Luvussa 6 kuvataan algoritmiliittymät ja luvussa 7 ulkoiset rajapinnat. Luku 8 käsittelee testausta, luku 9 rajoitteita ja luku 10 ohjelmiston muita ominaisuuksia.

2 Yleiskuvaus (TK)

`pics98`-projekti on PICSearch-ohjelmiston version 2 jatkokehitystä. Järjestelmän olemassa olevien toimintojen virheitä korjataan ja käyttöliittymä uusitaan. Uusia ominaisuuksia lisätään prioriteettijärjestyksessä (ks. aliluku 2.3 Muutokset).

2.1 Yleinen toiminta

Järjestelmä jakautuu kahteen toiminnalliseen osaan: asiakas- ja palvelinohjelmaan (ks. Kuva 1). Asiakasohjelman suorittamat tietokantaoperaatiot tapahtuvat palvelinohjelman tarjoamien palveluiden kautta. Tarkempi kuvaus järjestelmän toiminnasta löytyy edellisten projektien vaatimusmäärittelydokumenteista (ks. [3] ja [4]).



Kuva 1: PICSearchin asiakas-palvelinmalli.

2.2 Ympäristö

Ohjelmisto toteutetaan ja testataan Linux-ympäristössä ja se vaatii Tcl/Tk-kirjaston (versio 7.6/4.2), X windows-ikkunointijärjestelmän ja TCP/IP-protokollaa noudattavan verkkoyhteyden. Toteutuskielenä on C++.

Ympäristöriippuvainen koodi on rajattu tiettyihin moduuleihin ja sen määrä pidetään minimissä, joten ohjelmisto on siirrettävissä pienin muutoksin muihin yo. vaatimukset täyttäviin ympäristöihin, esim. eri UNIX-järjestelmiin tai Windows95/NT-ympäristöön.

2.3 Muutokset

Järjestelmään `pics98`-projektin aikana tehtävät parannukset on jaettu kolmeen prioriteettiluokkaan niiden tärkeyden perusteella: luokkaan 1 kuuluvat muutokset ovat tärkeimpiä, luokkaan 3 kuuluvat vähiten tärkeitä.

Prioriteetti 1

- Tietokantojen nimissä esiintyvien välilyöntien aiheuttaman virheellisen käyttäytymisen korjaaminen.
- ID- ja nimihakujen korjaaminen.
- Kuvan lisäyksen attribuuttivälityksen puutteellisuuksien korjaaminen.
- Vierityspalkin liikuttamisen aiheuttamien turhien kuvahakujen poistaminen.
- Vieryspalkin muuttaminen kuvien määrän mukaan skaalautuvaksi.
- Käyttöliittymän vaihtoehtoisten ikkunakokojen toteuttaminen.
- Anomyynikäyttäjän oikeuksien asettamisen korjaaminen..
- Peräkkäisiä hakuja suoritettaessa esiintyvän virheellisen 100% osuma -ilmoituksen korjaaminen.
- Monilisäyksen (Batch Insert) estäminen, jos tietokantaa ei ole valittu
- Monilisäyksen palautteen parantaminen.

Prioriteetti 2

- Tietokanta-/algoritmikohtaisen informaation tallettamisen ja esittämisen toteuttaminen.
- Tietokannan roskienkeruun toteuttaminen.
- Viimeksi käytettyjen palvelimien listan toteuttaminen.
- Kuvien säilyttämiseen tarkoitetun välimuistirakenteen toteuttaminen asiakasohjelmaan.
- Erillisen algoritmien lisäysohjelman toteuttaminen.
- Kuvankäsittelyohjelman (Image Editor) korjaaminen tai korvaaminen ulkoisella ohjelmistolla.

Prioriteetti 3

- Tietokantajärjestelmän toiminnan tehostaminen.
- Tietokantakonvertterin (olemassa olevien tietokantojen muuntaminen ohjelmallisesti uusimpaan muotoon) toteuttaminen.

3 Tietokuvaus (SK)

Tässä luvussa käydään läpi niitä tiedonsijainnin ja -muodon kannalta oleellisia asioita, joihin `pics98`-projektiryhmä tulee vaikuttamaan. Alkuperäisen PICSearch- ja viimeisimmän NetPICSearch- projektien toimesta on ohjelman käyttämät tietorakenteet ja tiedon välitystavat kehittyneet käyttökelpoisiksi. Tämän projektiryhmän tarkoituksena on ensisijaisesti tehostaa, korjata ja parantaa nykyisiä ohjelman ratkaisuja, ei lähteä muuttamaan ohjelman rakenetta.

3.1 Vaatimukset ja suorituskyky

Muutosten motivaationa on ollut tietokantahakujen tehostaminen ja siten ehdottomana vaatimuksena on, että `pics98`-projektiryhmän ohjelmistoon tekemät muutokset eivät hidasta tietokantahakuja tai ohjelman muuta toimintaa PICSearch v2:een nähden. Jos jokin toiminto hidastuu johtuen lisäntyneestä tiedonvälityksestä, täytyy menetetty tehokkuus saada takaisin muissa toiminnoissa siten, että ohjelmiston kokonaistehokkuus pysyy vähintäänkin ennallaan.

Tietokantamoottorissa ja tietoliikennemoduuleissa tehtävät muutokset eivät saa edellyttää käyttäjältä uusien toimenpiteiden oppimista, eli lisätehokkuuden saavuttaminen on rakennettava nykyisen rungon ympärille siten, että muutokset eivät näy järjestelmän käyttäjälle muuten kuin nopeutuneena toimintojen käsittelynä.

3.2 Tietokanta

Nykyistä tietokantaratkaisua tehostetaan. Ratkaisu, jossa tietokantamääritykset, algoritmit, kuvista lasketut tunnisteet (Signature), pienoiskuvat (Thumbnails), palvelimen asetukset ja käyttäjätiedot pidetään erillisissä tietokannoissa, pysyy samana. Näiden erillisten tiedostojen käytettävyyttä ja tehokkuutta parannetaan ohjelmiston tietokantamoottoria kehittämällä. Tietokannoille rakennetut indeksit säilytetään mutta niiden olemassaolo tehdään palvelimelle entistä merkittävämmäksi siten, että indeksitiedostoja käytetään entistä tehokkaammin osana tietokantahakuja.

3.3 Tietoliikenne

Aikaisempien projektiryhmien ohjenuoraa ”mitä vähemmän verkon yli siirretään, sitä parempi” noudatetaan myös `pics98`-projektiryhmässä. Ohjelmistoon tehtävät parannukset johtavat siihen, että osa toiminnoista tulee aiheuttamaan entistä enemmän liikennöintiä asiakkaan ja palvelimen välillä. Asiakkaan päähän toteutettava välimuisti vuorostaan vähentää huomattavasti liikennöintiä asiakkaan ja palvelimen välillä. Tarkoituksena on, että uusilla tehostamiskeinoilla kompensoidaan ohjelmiston kasvanutta tiedonsiirotarvetta.

3.4 Asetustiedostot

PICSearch v2 -ohjelmistossa on palvelimella asetustiedosto, jonka asetuksia voidaan muuttaa asiakasohjelmalla riippuen käyttäjän käyttöoikeuksista. Palvelimeen liittyviä asetuksia lisätään vain minimaalisesti. Asiakkaan päähän toteutetaan, nykyisin vielä puuttuva, asetustiedoston käyttö. Asiakkaan asetustiedostoon tallennetaan tietoja käyttöliittymästä, käytetyistä palvelimista ja tietokannoista, ajoympäristöstä ja käyttäjästä.

Oleellista asiakkaan pään asetustiedoston toteutuksessa on, että asiakasohjelmisto muistaa edellisen kerran ajoympäristön. Näin loppukäyttäjän ei tarvitse jokaisen käynnistyskerran yhteydessä muokata ajoympäristöä mieleisekseen.

Seuraavassa esitetään asiakkaan päähän toteutettavan asetustiedoston runko.

picsearch.cfg

```
[databases]
..
..
[gui]
..
..
[runtime]
..
..
[used servers]
..
..
[user]
..
..
```

4 Tiedonvälitys (SK)

PICSearch-ohjelmistoon tehtävät muutokset näkyvät tiedonvälityksessä sekä eri osien välillä, että osien itsensä sisällä. Jotkut lisäominaisuudet kasvattavat liikenettä, toiset vuorostaan vähentävät sitä, kuitenkin siten että kokonaisuudessa tiedonvälitys pienenee tai pysyy vähintään ennallaan PICSearch v2:een verrattuna. Uusien ominaisuuksien toteutus näkyy sekä asiakkaassa että palvelimessa, joita seuraavassa käsitellään.

4.1 Palvelimen sisäinen tiedonvälitys

Oleellisin muutos palvelimen sisäisessä tiedonvälityksessä on seurausta algoritmien lisäämiseen ja poistamiseen toteutettavasta algoritmien hallintaohjelmasta. Vain pääkäyttäjä (SuperUser) voi lisätä tai poistaa algoritmeja. Palvelimen täytyy huomata uuden algoritmin lisäys ja laskea kyseisellä algoritmilla valitulle tietokannalle kuvatunnisteet (Signature). Vastaavasti kun algoritmien hallintaohjelmistolla poistetaan joku algoritmi, täytyy palvelimen poistaa algoritmiin liittyvät kuvatunnisteet kuvatietokannasta. Palvelin laskee kuvatunnisteet aina myös asiakkaan suorittaman kyselyn yhteydessä. Nykyisessä versiossa käyttäjää ei informoida algoritmin lisäämisestä tai poistamisesta johtuvasta asiakkaan ja palvelimen uudelleenkäynnistämistarpeesta.

4.2 Asiakkaan sisäinen tiedonvälitys

Luvussa 3.3 kuvatun asiakkaan päähän toteutettavan lisämuistin ohella toteutetaan asiakkaaseen asetustiedoston käyttö. Asetustiedostoon tallennetaan ajoympäristön oleellisimmat parametrit, jotka noudetaan aina käynnistyksen yhteydessä.

4.3 Asiakkaan ja palvelimen välinen tiedonvälitys

Tiedonvälityksen parantaminen asiakkaan ja palvelimen välillä sekä algoritmien hallintaohjelmiston toteuttaminen ovat suurimmat `pics98`-projektiryhmälle asetetut haasteet. Asiakkaan ja palvelimen välisen tiedonvälityksen tehostaminen saavutetaan asiakkaan päähän toteutettavalla välimuistilla. Jatkossa algoritmit sijaitsevat vain ja ainoastaan palvelimella. Tämä muutos johtaa siihen, että asiakkaan suorittaman kyselyn yhteydessä täytyy koko kyselyn kohteena oleva kuva lähettää palvelimelle. Kuvan saavuttua kokonaisuudessaan palvelimelle, laskee palvelin kuvaan liittyvät tunnisteet (Signature) kullakin valitulla algoritmilla ja suorittaa vertailun tietokannassa oleviin vastaaviin tunnisteisiin algoritmeihin liittyvillä vertailuoperaatioilla. Tämä muutos aiheuttaa tiedonvälityksen hetkellisen kasvamisen PICSearch v2:een verrattuna, mutta kokonaistehokkuus pysyy vähintään yhtä hyvänä.

5 Käyttöliittymä (TK)

Pääikkunasta laaditaan kolme versioita eri resoluutioille: 800*600, 1024*768, 1280*1024.

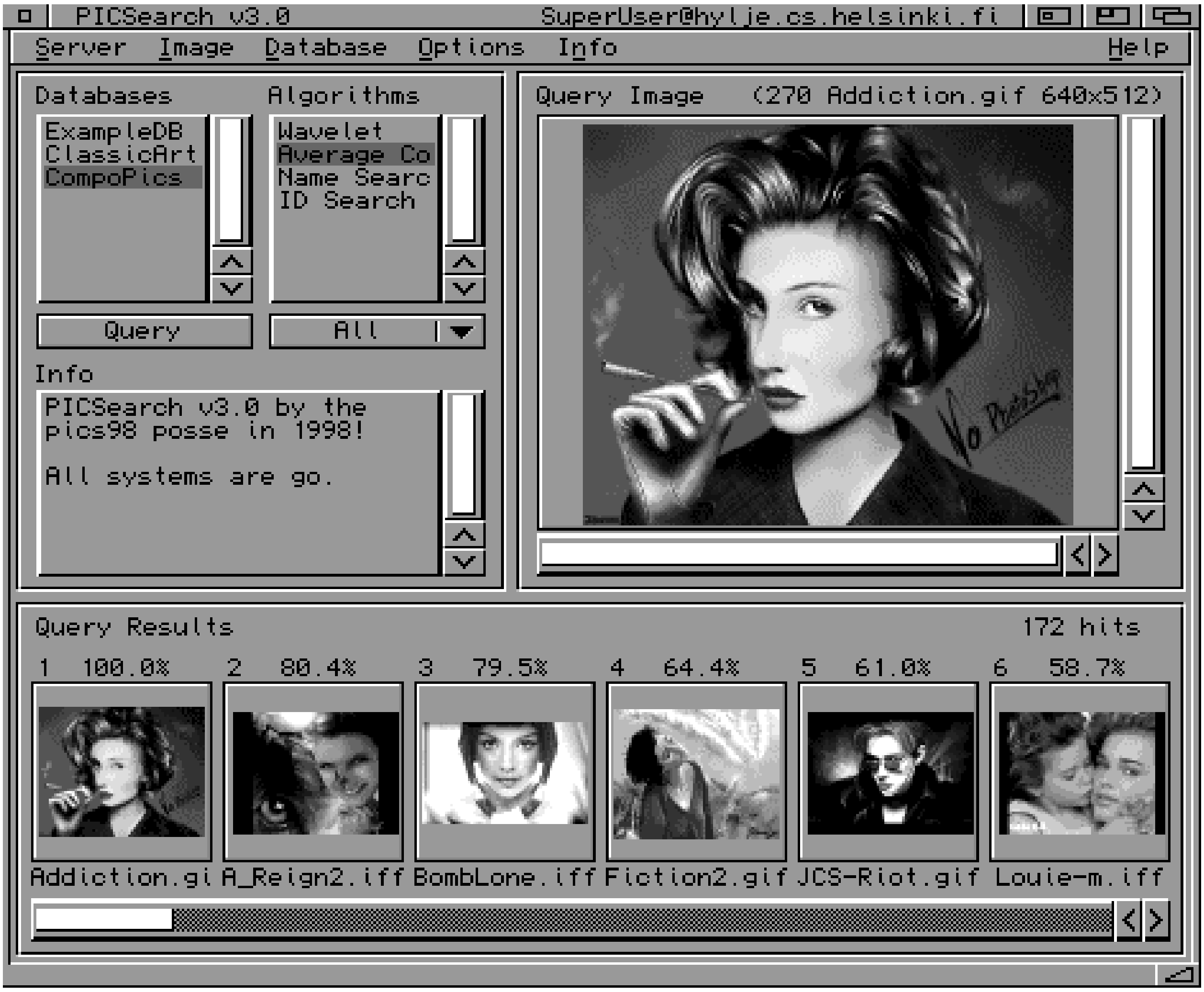
HUOM! Tässä luvussa esiintyvät ikkunoiden kuvat eivät välttämättä ole lopullisia versioita.

5.1 Muutokset ja uudet ominaisuudet

Pääikkuna

- Tuloskuvat sijoitetaan vaakatasoon pääikkunan alaosaan. Tuloskuvat ovat yhdellä tai useammalla rivillä riippuen valitusta ikkunakoosta. Tuloskuvien yhteydessä on niiden lukumäärä.
- *Databases*- ja *Algorithms*-listat sijoitetaan esimerkkikuvan viereen. *Images*-lista poistetaan ja vastaava toiminnallisuus toteutetaan valikoilla (ks. Taulukko 3). Myös *Get Thumbnail* ja *Get Image* -painikkeiden toiminnot siirretään tähän valikkoon.
- *Algorithms*-listan yhteyteen lisätään pudotusvalikko, jolla hakutyyppi voidaan valita ja *Search Type* -valikko poistetaan.
- *Ask Parameters* -valintanappi poistetaan ja toiminto siirretään *Client Options* -ikkunaan.
- Infoikkuna sijoitetaan *Databases*- ja *Algorithms*-listojen alle esimerkkikuvan viereen.
- *Image Editor* ja *Attributes* -painikkeet poistetaan ja ko. toiminnot siirretään valikoihin.
- Hakukuvan yhteyteen laitetaan näkyviin osa sen attribuuteista: id (jos on), tiedostonimi, leveys ja korkeus.
- Hakukuvasta tehdään skaalautuva, so. koko kuva näkyy aina hakukuvaikkunassa riippumatta kuva-alueen koosta. Skaalauksen voi kytkeä päälle tai pois asetuksista.

- Käytössä olevan palvelimen nimi siirretään *Server*-valikon yhteydestä ohjelman otsikkopalkkiin, jossa näkyy myös käyttäjätunnus.



Kuva 2: Käyttöliittymän pääikkuna

Valikot

Osa PICSearch v2:n pääikkunan toiminnallisuudesta siirretään valikoihin ja niiden sisältö organisoidaan uudelleen. Taulukoissa 1-6 on kuvattu järjestelmän valikot. Uudet toiminnot on merkitty vahvennettuina ja niiden toiminta on selvitetty kunkin taulukon oikeanpuoleisessa osassa.

Server...			
	Connect To...		
		Localhost	Ottaa yhteyden samassa työasemassa olevaan palvelimeen.
		Remote Server	Avaa palvelindialogi-ikkunan yhteyden muodostamiseksi etäpalvelimeen.
		—	
		Server 1	Pikavalinta 1.
		Server 2	Pikavalinta 2.
		...	
		Server n	Pikavalinta n.
	Disconnect		
	—		
	Database Maintenance		
	User Maintenance		
	—		
	Change Password		
	—		
	Quit		

Taulukko 1: Server-valikko

Database...		
	Insert Image Into Database	
	Delete Image From Database	
	—	
	Batch Insert	

Taulukko 2: Database-valikko.

Dialogi-ikkunat

- *Image Attributes* -ikkunassa editoitavista kentistä tehdään kiinteistä tekstikentistä eroavia. *Image Attributes* ja *Signatures* -ikkunat yhditetään.
- *Database Maintenance* ja *Algorithm Management* -ikkunat yhdistetään.
- *Query Progress* -ikkunaan lisätään kyselyn edistyminen prosentteina.
- *Change Password* -ikkunaan laitetaan näkyviin minkä tunnuksen salasanaa ollaan vaihtamassa.
- *Client Options*-ikkunaan lisätään mahdollisuus määritellä ulkoinen kuvankäsittelyohjelma. Ikkunakoon valinta siirretään *Options*-valikosta *Client Options*-

Image...		
Available Images...		Hakukuvan valinta. Toimii samoin kuin PICSearch v2:n *Images*-lista.
	Image 1	Hakukuva 1.
	Image 2	Hakukuva 2.
	...	
	Image n	Hakukuva n.
—		
New Query Image		
Open Query Image		
Close Query Image		
Save Query Image		
—		
Get Selected Thumbnail		
Get Selected Image		
—		
Scale Query Image		Kytkee hakukuvan skaalauksen päälle/pois.
Image Attributes		
Image Editor		

Taulukko 3: Image-valikko.

Options...	
Edit Client Options	
Edit Server Options	Sama kuin PICSearch v2:n *Server*-valikon *Options*.

Taulukko 4: Options-valikko.

Info...	
On Active Server	Sama kuin PICSearch v2:n *Server*-valikon *Info*
On Selected Database	Näyttää tietokantakohtaiset tiedot (kuvien lukumäärä, jne.) infoikkunassa.
On Selected Algorithm	Näyttää algoritmikohtaiset tiedot (algoritmin laatija, versio, jne.) infoikkunassa.
—	
On PICSearch	Versiotiedot, tekijät, jne.

Taulukko 5: Info-valikko.

Help... On PICSearch	Lyhyet käyttöohjeet.

Taulukko 6: Help-valikko.

ikkunaan.

- *Client Options* ja *Server Options* -ikkunoihin lisätään painikkeet asetusten tallentamista ja lataamista varten.

Toiminnot, joita käyttäja ei saa (käyttäjätunnuksensa oikeuksien puitteissa) tai ei voi suorittaa, ovat pääikkunassa, valikoissa ja dialogi-ikkunoissa harmaina, so. niitä ei voida valita.

6 Algoritmiliittymät (MP)

Algoritmiliittymät ovat algoritmien valintaan ja käsittelyyn liittyviä rajapintoja.

Asiakkaan käyttöliittymässä on lista palvelimen tarjoamista algoritmeistä. Tässä *Algorithms*-listassa näkyvät ainoastaa ne algoritmit, joita asiakas voi käyttää valitsemassaan tietokannassa. Algoritmilista päivittyy tietokantaa vaihdettaessa. Algoritmin valinnan yhteydessä infoikkunaan tulostuu tietoa algoritmista.

Algoritmien lisäystä ja poistamista helpotetaan tekemällä erityinen algoritmienkäsittelyohjelma. Kyseisellä ohjelmalla algoritmeja voidaan myös muokata.

7 Ulkoiset rajapinnat (MP)

PICSearch-ohjelmisto käyttää kuvan muokkaamiseen ulkoista kuvankäsittelyohjelmaa (Image Editor). PICSearch v2:n käyttämä piirto-ohjelma ei toimi. Kuvankäsittelyohjelmassa olevat virheet joko korjataan tai tilalle vaihdetaan uusi piirto-ohjelma.

8 Testaus (JL)

Aikaisemmat projektit eivät ole kiinnittäneet suuremmissa määrin huomiota PICSearch-ohjelmiston testaukseen. Tämänkertaisessa `pics98`-projektissa painopistettä siirretään ohjelmiston laajentamisesta testaukseen.

Tulosteet ja syötteet ovat käyttäjälle näkyvä osa ohjelmistosta. Testauksessa pyritään varmistamaan, etteivät mitkään syötteet kaada ohjelmaa. Oikeat syötteet hyväksytään ja vääristä annetaan virheilmoitus. Tarvittaessa syötteistä pyritään löytämään ekvivalenssiluokat. Näiden luokkien pohjalta pyritään varmistumaan ohjelman oikeasta toiminnasta.

Tulosteista tarkastetaan, että ne ovat kussakin tilanteessa järkeviä.

9 Rajoitteet (JL)

Käytettävän tietokonelaitteiston tulee olla riittävän tehokas. Vähimmäisvaatimuksena on 486-tasoinen prosessori ja 16–32 MB keskusmuistia.

Eri versiota olevat palvelin- ja asiakasohjelmat eivät ole keskenään yhteensopivia. PICSearch v2-palvelimen kuvatietokanta voidaan kuitenkin konvertoida erityisellä projektin tuottamalla konvertointiohjelmalla uudemmalla palvelinversiolla käytettäväksi.

10 Muut ominaisuudet (MP)

Järjestelmän keskeisimmät ei-toiminnalliset ominaisuudet ovat laajennettavuus ja ylläpidettävyys. Näitä ominaisuuksia toteutetaan olio-tekniikalla, hyvillä rajapintamäärittelyillä ja kattavalla dokumentaatiolla. Kolmas keskeinen ei-toiminnallinen ominaisuus on siirrettävyys. Siihen pyrkimiseksi sovellusalustasta riippuvaisia piirteitä vältetään. Siirrettävyys ja ylläpidettävyys pidetään samalla tasolla kuin NetPICSearch-projektilla (ks. [4]). Järjestelmän suorituskyky ei ole ensisijainen tavoite. Suorituskyky suhteutetaan kuvatietokannan kokoon.

Viitteet

[1] **K. Lemström, J. Korte, P. Kuusi, P. Kyheröinen, and P. Päiväkumpu:**
PICSearch - A Platform for Image Content-based Searching Algorithms, (Technical Report C-1997-48), Proceedings of the Sixth International Conference in Central Europe on Computer Graphics and Visualisation 98 (WSCG'98), edited by Vaclav Skala, Plzen, Czech Republic, February 9-13, 1998, pp. 222-229.
`http://www.cs.helsinki.fi/~klemstro/publications.html`

[2] **K. Lemström, P. Kuusi, P. Kanerva, S. Koponen, S. Männynsalo, J. Riihimäki, J. Teinilä, T. Virtanen:**
A Client-Server Tool for Image Retrieval Applications, The Challenge of Image Retrieval, edited by John P. Eakins, David J. Harper and Joemon Jose, Newcastle upon Tyne, United Kingdom, February 5-6, 1998.
`http://www.cs.helsinki.fi/~klemstro/publications.html`

[3] **J. Korte, P. Kuusi, P. Kyheröinen, P. Päiväkumpu:**
PICSearch-projektin (PICSearch v1) vaatimusmäärittely, Teknillinen korkeakoulu, 1996.
`http://www.hut.fi/ pkuusi/oht/documents/tm.html`

[4] **P. Kanerva, S. Koponen, S. Männynsalo, J. Riihimäki, J. Teinilä, T. Virtanen:**
NetPICSearch-projektin (PICSearch v2) vaatimusmäärittely, Helsingin yliopisto, Tietojenkäsittelytieteen laitos, 1997.
`http://www.cs.Helsinki.FI/group/netpics/VM/VaatM12.html`

Liite: Käytetty erikoissanasto

Attribuutti (Attribute) Kuvaan liittyvä tieto, esim. nimi, korkeus, leveys.
Hakualgoritmi (Search Algorithm) Algoritmi, joka laskee kuvatunnuksen kuvasta ja määrittää kuinka hyvin kaksi kuvatunnusta vastaavat toisiaan.
Hakukuva (Query Image) Käyttäjän valitsema kuva, jonka perusteella kuvahaku suoritetaan.
ID (Identification) Kuvalle annettu yksiselitteinen numeerinen tunniste.
Kuvankäsittelyohjelma (Image Editor) Ohjelma, jolla kuvaa voi manipuloida.
Kuvatunnus (Signature) Kuvasta laskettu kompakti esitys, joka pyrkii tiiviisti kuvaamaan kuvan tai osakuvan sisältöä.
Monilisäys (Batch Insert) Operaatio, joka lisää tietokantaan useamman kuvan kerralla.
Pienoiskuva (Thumbnail) Kuvasta skaalaamalla muodostettu pienennys.
Tuloskuvat (Result Images) Kuvahaun perusteella löydetyt kuvat.