

Kalervo Järvelin ja
Timo Niemi

Hajautettujen faktatietokantojen käytön yksinkertaistaminen

Järvelin, Kalervo and Niemi, Timo, Hajautettujen faktatietokantojen käytön yksinkertaistaminen [Simplifying Retrieval from Distributed Fact Databases]. Kirjastotiede ja informatiikka 9 (1): 3–16, 1990.

The problems of fact retrieval from modern distributed and heterogenous fact databases are first discussed. In order to simplify retrieval, an intermediary system for fact retrieval, analogous to those for bibliographic retrieval, is needed. An imaginary example of a distributed fact retrieval process, based on current tools and systems, is analyzed in order to clarify the requirements of such an intermediary system. The retrieval process as supported by an intermediary system for fact retrieval is then described and the properties of such a system discussed. A research project aiming at fact simplification of retrieval is introduced.

*Addresses: Kalervo Järvelin, University of Tampere, Department of Library and Information Science, P.O.Box 607, SF-33101 Tampere, Finland.
Timo Niemi, University of Tampere, Department of Computer Science, P.O.Box 607, SF-33101 Tampere, Finland.*

1. Johdanto

Merkittävä osa tieteellisen tutkimuksen, julkisen hallinnon ja talouselämän tarvitsemista tiedoista talletetaan nykyisin tietokannoissa. Tämä kehityskulku alkoi pääasiassa 1960-luvulla ja jatkuu yhä kiihtyvällä vauhdilla. Tämä on 1970-luvulta lähtien Suomessakin johtanut tietokantojen julkiseen, ja usein kaupalliseen, tarjontaan kaikkien halukkaiden (ja maksukykyisten) käyttöön. Tietokantojen käytön laajuuteen vaikuttaa niiden käytön kustannusten ohella olennaisesti niiden käytön helppous. Tiedonhaun tutkimuksessa on tällä vuosikymmenellä pyritty luomaan järjestelmiä ja menetelmiä tukemaan loppukäyttäjien (tiedontarvitsijoiden) itsenäisesti tekemää tiedonhakua, jossa ei siis tarvita välittäjän (kirjastonhoitajan tai informaattikon) apua [esim. [23] [25]]. Tämä on

johtanut mm. vilkkaaseen tiedonhaun välittäjäjärjestelmien tutkimukseen (lähinnä viitetietokantoja varten, esim. [6] [10] [15] [12] [26]). Toisaalta kuitenkin monilla eri tahoilla ja monenlaisten tietokantojen käyttöyhteyksissä on esitetty näkemyksiä, että tietokantojen käyttö on usein liian vaikeaa joko ammattimaisille välittäjille tai ainakin loppukäyttäjille (edellisten ohella esim. [3] [11] [21] [30] [33]).

Mielestämme tiedon tallennuksen ja haun tutkimuksen päätavoite on kehittää käsitteitä, menetelmiä ja järjestelmiä, joiden avulla kaikki tieto, olipa se missä tahansa muodossa ja missä tahansa paikassa, saadaan vaivattomasti kenen tahansa sitä tarvitsevan ulottuville ja siten esitettynä, että tarvitsijan on mahdollisimman helppo tämä tieto omaksua. Tämä näkemys laajentaa oppikirjoissa (esim. [17] [28]) esitettyjä näkemyksiä samaan suuntaan, mihin käytännön tiedonhakutoiminta on kehittynyt tällä vuosikymmenellä.

Tässä artikkelissa tarkastellaan ongelmaa, *miten voidaan yksinkertaistaa hajautettujen, heterogeenisten faktatietokantojen käyttöä*. Faktatietokantojen käyttöön liittyy useita sellaisia ongelmia, joita muiden tietokantojen käyttöön ei liity ja joita on tutkittu riittämättömästi. Faktatietokannat ovat nykyisin tärkeitä tiedonhankinnan välineitä, joiden käyttöön tai käyttämättömyyteen liittyy myös yhteiskunnallisia kysymyksiä (esim. kilpailukyky ja tasa-arvo). Lisäksi ne muodostavat varsin uuden alueen kirjastotieteen ja informatiikan tutkimuksessa. Tiedonhaun yksinkertaistamiseen voidaan pyrkiä ns. faktahaun välittäjäjärjestelmien kehittämällä. Tällaisen järjestelmän esikuvana ovat viitteenhaun välittäjäjärjestelmät ja sen tehtävänä on koota ja korvata monen sisällöltään vain osittain yhteensopivan tietokannan, monen tiedonhallintajärjestelmän, monen tietomallin ja monen tietoliikenneverkon samanaikaisessa käytössä tarvittava ihmisten asiantuntemus ja tarjota se tietokantojen käyttäjien käytettäväksi automaattisena välineenä.

Tässä artikkelissa jäsennetään tietokantoja ja niihin liittyviä keskeisiä käsitteitä ja faktatietokantojen nykyaikaista käyttöympäristöä kuvaavia tämän tutkimuksen kannalta merkityksellisiä piirteitä. Seuraavaksi tarkastellaan yksinkertaistettua esimerkkiä niistä ongelmista, joita hajautettujen faktatietokantojen käyttäjät nykyisin kohtaavat. Tällaisten ongelmien automaattinen ja/tai puoliautomaattinen ratkaiseminen on faktahaun välittäjäjärjestelmän tehtävä. Tämän jälkeen tarkastellaan faktahaun välittäjäjärjestelmän käyttöliittymää ja sen tukemaa tiedonhakuprosessia. Lopuksi esitellään faktahaun välittäjäjärjestelmiin kohdistuva tutkimushanke.

2. Nykyaikainen tietokantaympäristö

2.1. Tietokantojen jäsennyksiä

Tietokantoja voidaan jäsentää kahden toisiaan täydentävän näkökulman kautta: tietojenkäsittelyopin kannalta, jolloin keskeisiä jäsennystekijöitä ovat tietokantojen rakenne ja tietojen käsittelyominaisuudet, sekä tiedonhakutoiminnan näkökulmasta, jolloin keskeisiä jäsennystekijöitä ovat tietokantojen tietosisältöön liitty-

vät tekijät.

Ryhtymättä yksityiskohtaiseen ja täsmälliseen määrittelyyn voidaan todeta, että *tietojenkäsittelyopin näkökulmasta* tietokannat jäsenyivät useisiin tyyppisiin käytettyjen tietomallien mukaan (esim. [34] [28]): hierarkkinen tietomalli, relaatiotietomalli, verkkotietomalli, oliotietomalli ja rekursiivinen tietomalli. Tietomalli määrittelee millaisia rakenteellisia alkioita tietokantaan voi sisältyä ja minkä tyyppisiä yhteyksiä niiden välillä voidaan muodostaa. Tietokanta koostuu aina jonkin tietomallin rakeneperiaatteiden mukaan jäsennetyistä tietoalkioista. Kunkin tietokannan käyttö tapahtuu *tiedonhallintajärjestelmän* (database management system, DBMS) avulla, joka määrää tietokannan tietomallin. *Hierarkkisessa tietomallissa* (hierarchical data model), jota kutsutaan myös NF²-tietomalliksi (*non-first normal form*), on rakenteen pääpiirre hierarkia. *Relaatiotietomallissa* (relational data model) kaikki tieto esitetään matemaattiseen relaation käsitteeseen perustuvina kaksikulotteisina taulukkoina. *Verkkotietomallissa* (network data model) tieto jäsenyy tietuetyyppeihin ja näiden välisiin yhteystyyppisiin, jotka muodostavat verkoston. Nämä kolme tietomallia (hierarkkinen, relaatio- ja verkkotietomalli) ovat perinteiset »suuret» tietomallit.

Edellisiin perustuvissa tietokannanhallintajärjestelmissä ei voida helposti suorittaa esim. kyselyä työntekijän kaikista esimiehistä (kaikilla hierarkkiatasoilla). Siksi 1980-luvulla on kehitetty voimakkaasti *oliotietomalliin* (object-oriented data model) ja *rekursiiviseen tietomalliin* (recursive data model) perustuvia tiedonhallintajärjestelmiä. Oliotietomallissa sallitaan rekenteellisesti monimutkaiset oliotyypit (kuten hierarkkiset rakenteet) ja niiden *kotelointi* (encapsulation) ts. monimutkaisten oliotyyppi-kohtaisten käsittelysääntöjen määrittely. Rekursiiviseen tietomalliin perustuva tiedonhallintajärjestelmä tarjoaa tavallisesti sekä perinteisen tiedonhallintajärjestelmän käyttöominaisuudet että tavallisesti matemaattiseen logiikkaan perustuvan kielen kaiken tarvittavan käsittelyn määrittelyyn. [34] [35] [36]

Nykyaikaisen *tiedonhakutoiminnan kannalta* tietokannat voidaan jäsentää moneen eri ryhmään. Yksi erotteluperuste on tietokannan tietojen suoruus vs. välillisuus: *Lähdetietokannoista* löytyy itse tieto (esim. haettu vastaus) ja *korviketietokannoista* löytyy tiedon korvike, joka

edustaa varsinaista tietoa ja jonka avulla se voidaan paikantaa ja etsiä käyttöön (esim. kirjallisuusviite tai konsultin kontaktitiedot). Faktatietokannat ovat yksi lähdetietokantojen pääryhmistä. Tilastotietoja julkisesti tarjolla olevien tietokantojen määrän kasvusta saa monista julkaisuista (esim. [13] [14] [19]). Pelkästään numeerisiksi tietokannoiksi luokitettavia faktatietokantoja on julkisesti tarjolla yli 1000 kpl ympäri maailmaa [14].

Faktatietokantojen olennaisia ominaisuuksia ovat tietojen strukturoituneisuus (tieto ja sen netään erikseen nimetyiksi tietoalkioiksi, jotka talletetaan kukin omissa kentässään) ja faktuaalisuus (tieto koskee tosiasioita ja esitetään tietoalkioina) sekä täsmäytyksen ehdottomuus (kysymyksen oikea esitystapa voidaan määrittellä etukäteen ja vastaukseksi saadaan kaikki tietokannassa oleva relevantti tieto). Kaikki nämä ominaisuudet poikkeavat tiedonhakututkimuksessa perinteisesti tarkasteltujen viitetietokantojen ominaisuuksista. Ne poikkeavat myös kaikkien niiden tietokantojen ominaisuuksista, joissa haku perustuu tekstiin (esim. tekstitietokannat).

2.2. Tietokantaympäristö

Nyky aikaista kaupallista faktatietokantaympäristöä voidaan luonnehtia seuraavilla piirteillä

- Laajuus: Erilaisia faktatietokantoja on tarjolla tuhansia
- Hajautus: faktatietokannat sisältyvät useisiin eri tietokone- ja tietoliikenneverkkoihin
- Valvonnan puute: tietokantojen sisällöstä ei ole ylikansallista valvontaa
- Valvonnan mahdottomuus: on epätodennäköistä, että valvontamekanismi voitaisiin rakentaa
- Tietojen nimennän epäyhtenäisyys eri tietokannoissa
- Tietojen yhteiskäytön vaikeus: tietojen epäyhtenäisyys edellyttää usein niiden yhteensovittamista
- Heterogeisuus: tietokantojen tietorakenteet, tietomallit ja tiedonhallintajärjestelmät vaihtelevat
- Monimutkaisuus: tietokantojen kyselykieleet ovat usein varsin mutkikkaita

Taulukko 2.1. Faktatietokantaympäristön ominaisuuksia.

(Taulukko 2.1), jotka vaikuttavat tiedonhaun helppouteen. Näitä piirteitä on tarkasteltu yksityiskohtaisesti tekijöiden toisessa raportissa (ks. [18]).

Tietojen epäyhtenäinen nimentä tarkoittaa, että läheisesti toisiinsa liittyviä tietoja on talletettu eri tietokantoihin täysin eri nimin ja keskenään epäyhtenäisiä nimentäperiaatteita noudattaen. Siksi on vaikea selvittää luotettavasti, mitkä tietoalkiot vastaavat merkitykseltään toisiaan. Usein tietokannat sisältävät esitystavaltaan epäyhtenäisiä, mutta periaatteessa yhteensopivia tietoja. Epäyhtenäisyys (inconsistency) voi koskea (a) tiedon teknistä esitysmuotoa; (b) tiedon esittämisen mittayksikköä (esim. \$ tai £; ISO:n mittayksiköt vs. unssit, gallonat, tuumat ja jaardit); (c) tietoalkoiden arvojen nimentää (esim. viikonpäivät eri kielillä); (d) tiedon mitaustapaa ja -tarkkuutta (esim. kuukausi-, neljännesvuosi- tai vuositaso tiedot taikka mitauksen (esim. indeksin) vaihteleva perusvuosi). Varsin usein on kuitenkin esitettävissä menetelmä, jolla epäyhtenäisyydestä johtuva näennäinen yhteensopimattomuus (incompatibility) voidaan torjua.

Nämä faktatietokantaympäristön ominaisuudet aiheuttavat kuitenkin mm. seuraavina kysymyksinä heijastuvia hankaluuksia faktatietokantojen käyttäjille (Taulukko 2.2).

- Onko käyttäjälle sopivaa tietoa olemassa tietokannoissa?
- Onko sellaisen tiedon hankinta suotavaa, välttämätöntä vai turhaa?
- Mistä muualta vastaavia tietoja on saatavilla?
- Missä kaikissa tietokannoissa tietoa on?
- Tuleeko hakea tietoa useista eri tietokannoista?
- Miten kuhunkin tietokantaan saa yhteyden?
- Miten tarvittu tiedot on nimetty ja esitetty eri tietokannoissa?
- Voidaanko kysely suorittaa eri tietokannoissa ja mitä se edellyttää?
- Miten kysely suoritetaan eri tietokannoissa?
- Miten tiedot eri lähteistä sovitetaan toisiinsa?

Taulukko 2.2. Faktatietokantojen käyttäjien ongelmia.

3. Esimerkki hajautetusta faktahausta

Hajautetun faktahaun ongelmien havainnollistamiseksi tarkastellaan tässä luvussa konkreettista, joskin kuvitteellista, esimerkkitilannetta, jossa nykyisten järjestelmien ja välineiden avulla haetaan tietoja partavesien hinnoista Euroopan eri maissa viime vuonna. Luonnollisella kielellä esitetty tiedontarve olkoon seuraava: »Anna niiden tuotteiden tavaramerkki, hinta ja myyntimaa, jotka ovat partavesiä, joita myydään Euroopassa ja joiden myyntiajan kohta on vuosi 1989».

Kuvitellaan hakutehtävää varten edelleen, että partavesiä (ja muitakin tuotteita) koskevat tiedot ovat saatavissa Lontoossa, Frankfurtissa ja Pariisissa sijaitsevista tietokannoista siten, että kussakin tietokannassa on joitakin, muttei välttämättä kaikkia Euroopan maita koskevia tietoja. Siis kokonaisinformaatio on jaettu maantieteellisesti: jokainen solmu sisältää partavesien kaikkia kiinnostavia ominaisuuksia koskevat tiedot jostakin Euroopan osasta. Päälekkäistä informaatiota tietokantojen kesken saattaa esiintyä. Eri paikkojen tietokantoja, niiden sisältöä ja rakennetta havainnollistetaan kuvassa 3.1.

Kuvassa 3.1. oletetaan yksinkertaisuuden vuoksi, että kaikki solmut sisältävät relaatiomalliin perustuvan tietokannan. Kuvasta kuitenkin ilmenee, että eri solmujen tietokannat ovat rakenteellisesti erilaisia: Pariisissa kaikki tiedot talletetaan yhdessä relaatioissa (taulukossa), Lontoossa kahdessa ja Frankfurtissa kolmessa. Pariisin tietokanta on ns. ensimmäisessä normaalimuodossa (first normal form, 1NF), Lontoon toisessa (2NF) ja Frankfurtin kolmannessa normaalimuodossa (3NF) (normaalimuodot, ks. [22] [31] [34]). Relaatioiden nimet vaihtelevat, samoin niiden attribuuttien nimet (taulukoiden sarakkeiden otsikot). Lisäksi toisiaan semanttisesti vastaavien attribuuttien arvot ovat erilaisia: Attribuutit *TYPE* (Lontoo), *TYPE* (Pariisi) ja *TYP (DIN)* vastaavat toisiaan, mutta Lontoossa arvot ovat kirjain-numero-koodeja, Pariisissa merkkijonoja ja Frankfurtissa numeroita. Voimme lisäksi olettaa (tämä ei näy kuvasta), että vastintribuuttien *COUNTRY* (Lontoo), *ETAT* (Pariisi) ja *STAAT* arvot ovat Lontoossa englanninkielisiä nimiä (esim. Spain, United Kingdom, Italy, Germany (W)), Pariisissa ranskankielisiä nimiä (esim.

kuten kuvassa 3.1.) ja Frankfurtissa standardinmukaisia lyhenteitä (esim. D, GB, F, I, E, DK, S, SF). Vielä voimme olettaa (tämäkään ei näy kuvasta), että vastintribuuttien *TIME* (Lontoo), *ANNEE* (Pariisi) ja *JAHRE* (Frankfurt) arvot ovat Lontoossa päiväkohtaisia (esim. 890531 = toukokuun 31. pv 1989), mutta Pariisissa ja Frankfurtissa vuosilukuja.

Nykyisin kyselyn suorittaminen edellyttäisi mm. seuraavia osatehtäviä. Yksinkertaisuuden vuoksi jätetään tässä huomiotta tietoliikenneverkon käytön edellyttämät toimenpiteet.

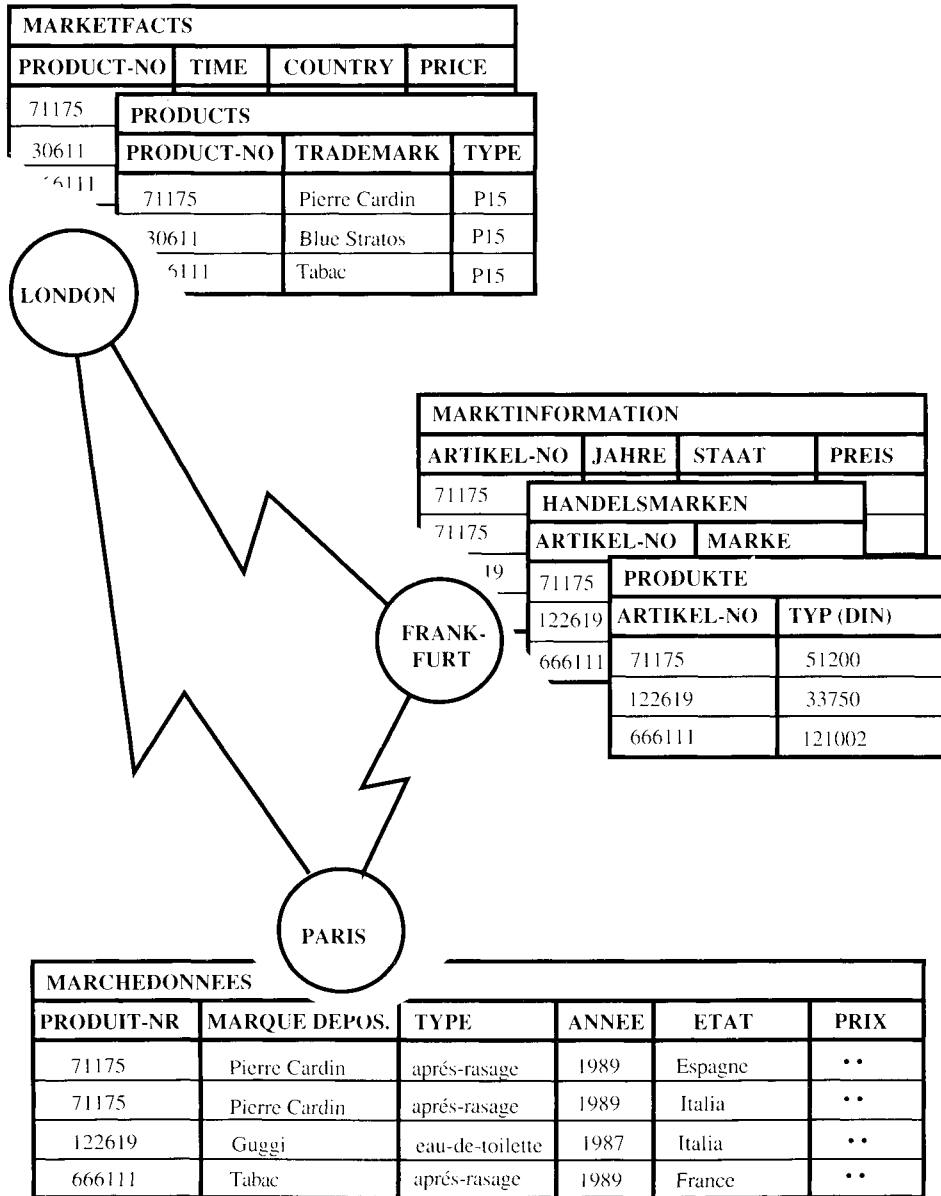
- Todetaan, että kysely kohdistuu tuotetietoihin ja että tuotteiden kiinnostavat ominaisuudet ovat *myyntihinta*, *myyntimaa*, *myyntiaika*, *tavaramerkki* ja *tuotetyyppi*. Rajoitteena edellytetään, että myyntimaa sijaitsee Euroopassa, myyntiaikatieto on vuodelta 1989 ja tuotetyyppi on partavesi.

- Kysely jaetaan osakyselyihin, joita tulee yksi kutakin solmua varten. Tässä tapauksessa koko kysely suoritetaan jokaisessa solmussa, koska jokainen solmu sisältää tietoa kaikista tuoteolioiden kiinnostavista ominaisuuksista.

- Kunkin solmun osalta selvitetään, missä relaatioissa partavesiä ja niiden ominaisuuksia koskevia tietoja on ja minkä nimiset attribuutit niihin liittyvät. Nämä tiedot ovat: (1) Lontoossa relaatio *MARKETFACTS* ja attribuutit *TIME*, *COUNTRY* ja *PRICE* sekä relaatio *PRODUCTS* ja sen attribuutit *TRADEMARK* ja *TYPE*; (2) Pariisissa relaation *MARCHEDONNEES* ja sen attribuutit *MARQUE DEPOSEE*, *TYPE*, *ANNEE*, *ETAT* ja *PRIX*; ja (3) Frankfurtissa relaatiot *PRODUKTE*, *HANDELSMARKEN* ja *MARKTINFORMATION* sekä niiden attribuutit (vastaavasti) (a) *TYP (DIN)*, (b) *MARKE*, ja (c) *JAHRE*, *STAAT* sekä *PREIS*.

- Todetaan kunkin kyselyssä käsiteltävän attribuutin tyyppi ja mahdolliset arvot. Tämän jälkeen selvitetään, miten kyselyn myyntimaa, myyntiaikaa ja tuotetyyppiä koskevat rajoitteet esitetään kunkin attribuutin määrittelyn edellyttämällä tavalla. Esim. todetaan, että kaikissa tietokannoissa tulee kaikki Euroopan maat luetella erikseen, kussakin solmussa eri tavalla. Lisäksi havaitaan, että Lontoossa myyntiaikatieto tulee esittää muodossa $TIME \geq 890101$ and $TIME \leq 891231$, ja muualla muodossa *ANNEE* = 1989 (tai *JAHRE* = 1989).

- Muodostetaan kussakin solmussa suoritettava kysely solmun tiedonhallintajärjestelmän kyselykielellä. Koska yksinkertaistetussa esi-



Kuva 3.1. Esimerkki hajautetusta ympäristöstä.

merkissämme kaikki solmut ovat relaationaalisia, voidaan odottaa, että kyselyt voidaan esittää SQL-tyyppisellä kyselykielellä (ks. kuva 3.2. ja sen selostus).

- Suoritetaan kysely kussakin solmussa. Saatavista mahdollisista vastauksista on esimerkiksi kuvassa 3.3.
- Vastauksia ei voida sellaisenaan yhdistää,

koska maa- ja hintatiedot eivät ole yhtenäisiä. Nämä tiedot muunnetaan yhteensopiviksi käännösmekanismeilla, joka tunnistaa eri maiden erikieliset nimet ja niiden lyhenteet ja kääntää ne esim. englanniksi sekä muuntaa hintatiedot sopivilla valuuttakurssikertoimilla esim. punniksi.

- Eri solmuissa saadut ja yhteismitallisiksi tehdyt tulokset lähetetään johonkin solmuun yhdistettäväksi ja lopuksi kyselyn esittäjälle. Lopputulos voisi olla kuvan 3.4. mukainen.

Kuvassa 3.2. esitetään SQL-tyyppiset kyselyt eri solmuja varten. Kyselyn perusrakenne SQL-kieleessä on aina *SELECT* <attribuuttiluettelo> *FROM* <relaatioluettelo> *WHERE* <ehtoluettelo>. Kuvan 3.2. kyselyjen *SELECT*-osissa ilmaistaan kunkin solmun tietokannan edellyttämällä tavalla, että halutaan *myyntihintaa*, *myyntimaata* ja *tavaramerkkiä* vastaavien attribuuttien arvot tulokseen. Kyselyjen *FROM*-osissa ilmaistaan kunkin solmun tietokannan ne relaatiot, joihin *SELECT*-osissa ilmaistut attribuutit sisältyvät.

(a) Pariisin kysely

```

SELECT MARQUE DEPOSEE, PRIX, ETAT
FROM MARCHEDONNEES
WHERE MARCHEDONNEES.TYPE = après-rasage
      AND MARCHEDONNEES.ANNEE = 1989
      AND MARCHEDONNEES.ETAT IN (France, Allemagne,
                                   Angleterre, Espagne, Italie, ...)

```

(b) Lontoon kysely

THIS PARAGRAPH COULD NOT BE READ IN FROM DISK

```

FROM MARKETFACTS.PRODUCTS
WHERE MARKETFACTS.TIME ≥ 890101
      AND MARKETFACTS.TIME ≤ 891231
      AND MARKETFACTS.COUNTRY IN (England, Wales, Scotland, Germany,
                                     Denmark, Ireland, Norway, ...)
      AND PRODUCTS.TYPE = P15
      AND MARKETFACTS.PRODUCT-NO = PRODUCTS.PRODUCT-NO

```

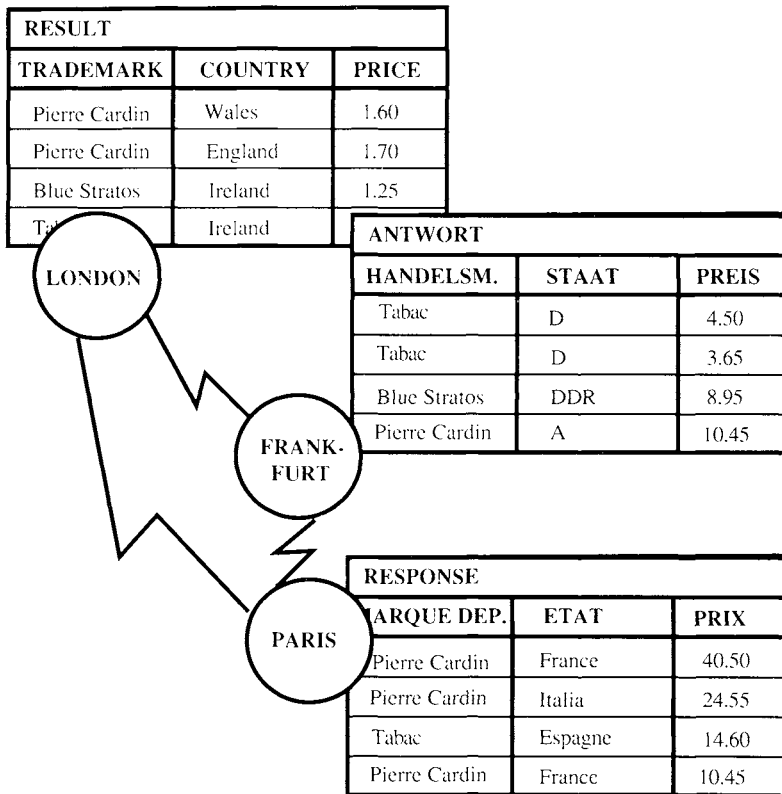
(b) Frankfurtin kysely

```

SELECT MARKE, PREIS, STAAT
FROM MARKTINFORMATION, PRODUKTE, HANDELSMARKEN
WHERE PRODUKTE.TYP(DIN) = 51200
      AND MARKTINFORMATION.JAHRE = 1989
      AND MARKTINFORMATION.STAAT IN (D, NL, B, S, SF, DDR, A, L, ...)
      AND MARKTINFORMATION.ARTIKEL-NO =
      HANDELSMARKEN.ARTIKEL-NO
      AND MARKTINFORMATION.ARTIKEL-NO = PRODUKTE.ARTIKEL-NO

```

Kuva 3.2. SQL-tyyppiset kyselyt eri solmuissa.



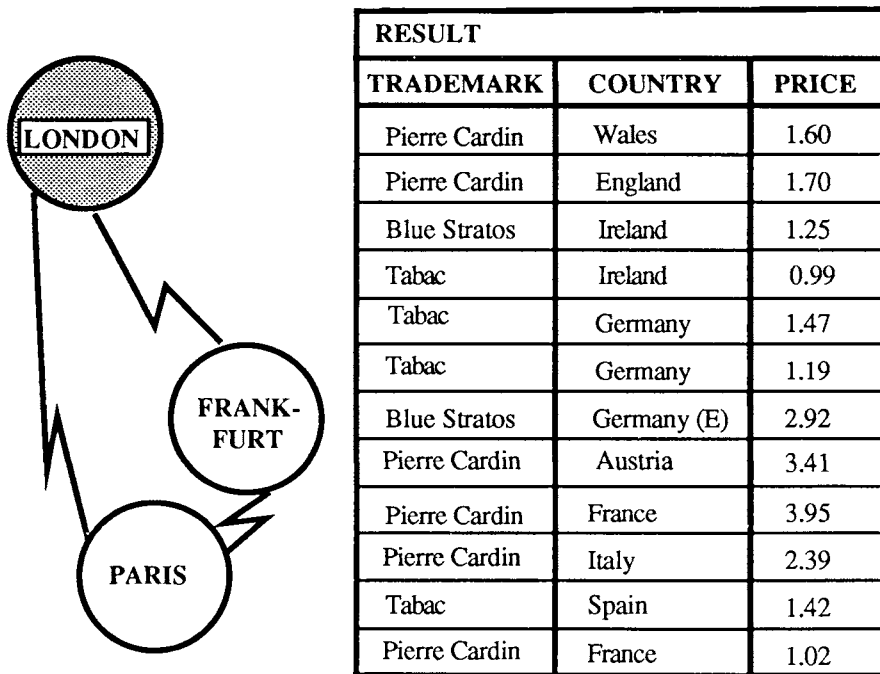
Kuva 3.3. Kyselyjen tulosesimerkkejä eri solmuissa.

Kyselyjen *WHERE*-osissa ilmaistaan ensiksi kunkin solmun tietokannan edellyttämällä tavalla, että *myyntimaata*, *myyntiaikaa* ja *tuotetyyppejä* vastaavien attribuuttien arvojen tulee olla vastaavasti jokin Euroopan maa, vuoden 1989 aika ja partavesi. Lisäksi kyselyjen *WHERE*-osissa ilmaistaan se, miten usean relaation tiedot voidaan liittää yhteen (ns. relaatioiden liittosehto). Tämä on esimerkissämme varsin yksinkertaista, koska Lontoossa ja Frankfurtissa kummassakin toimii tuotenumero eri relaatioita yksikäsitteisesti yhdistävänä tekijänä. Näin ollen Lontoossa liittosehto on *MARKETFACTS.PRODUCT-NO=PRODUCTS.PRODUCT-NO*, joka kertoo että relaation *MARKETFACTS* attribuutin *PRODUCT-NO* arvon tulee olla sama kuin relaation *PRODUCTS* attribuutin *PRODUCT-NO* arvo. Mitä useammassa relaatiossa tarvittavat tiedot sijaitsevat, sitä monimutkai-

sempi liitos (l. sitä useampi liittosehto) kyselyssä tarvitaan. Vaikka kaikissa solmuissa onkin olemassa toisiaan vastaavat attribuutit, on Frankfurtin kysely muita monimutkaisempi, koska siinä tarvitaan kolmea eri relaatiota.

Eri solmuissa suoritetuista kyselyistä saadusta tuloksista on esimerkki kuvassa 3.3. Attribuuttien *TRADEMARK*, *HANDELSMARKE*, ja *MARQUE DEPOSEE* arvot ovat yhtenäisiä, mutta muiden vastinattribuuttien arvot eivät ole. Maiden nimet ovat eri kielisiä ja hintatiedot ovat solmujen paikallisissa valuutoissa. Ennen tietojen yhdistämistä on niiden tekeminen yhteismitalliseksi välttämätöntä.

Kuvassa 3.4. eri solmujen tulokset on muutettu yhtenäisiksi käyttäen Lontoon solmun nimentöjä ja mittayksiköjä, lähetetty Lontoon ja yhdistetty siellä yhteiseksi kyselyn tulorelaatioksi.



Kuva 3.4. Kyselyn eri solmuista yhdistetty lopputulos.

4. Faktahaun välittäjäjärjestelmä

Faktahaun käytännön ongelmien ratkaisemisen edellyttää järjestelmää, joka tukee läpinäkyvää (transparent) tiedonhakua. Tällöin tiedonhakijan ei tarvitse huolehtia hakujen suorittamisen teknisistä yksityiskohdista (mitenkysymykset), vaan hän voi keskittyä määrittämään ja kuvaamaan tiedontarpeensa (mitäkysymykset). Järjestelmän tehtäväksi jää selvittää, mistä tiedot ovat saatavilla, miten yhteys tietokantoihin muodostetaan, kuinka kyselyt esitetään, miten vastaukset sovitetaan toisiinsa jne. Tällaista järjestelmää kutsumme *faktahaun välittäjäjärjestelmäksi* (intermediary system for fact retrieval) esikuvansa kirjallisuusviitteiden hakuun tarkoitettuja välittäjäjärjestelmiä mukailien (esim. [6] [10] [15] [12] [26]). Tarkastelemme seuraavassa faktahaun välittäjäjärjestelmän käyttöliittymää, sen tukemaa hakuprosessia, ja sen tehtäväalueita.

4.1. Käyttöliittymä

Faktahaun välittäjäjärjestelmän käyttöliittymän jäsentämisessä ovat käsitteet olio, olion ominaisuus ja olioiden suhde keskeisiä. Oliolla tarkoitetaan mitä tahansa reaalisen tai abstraktin maailman tiedon kohdetta, joka voidaan nimetä, jota voidaan tarkastella ja josta voidaan tallettaa tietoa — esim. tuote (reaalinen) tai muuttuja (abstrakti). Olion ominaisuudet jäsentävät olioita koskevan tiedon — esim. tuotteen hinta, paino, väri tai muuttujan perusjoukko ja arvojen vaihteluväli. Olioiden suhdetiedot kuvaavat olioiden konkreetteja tai abstrakteja keskinäisiä sijainteja, vuorovaikutusta, jne. — esim. tuotteen valmistaja- tai myyntimaatieto tai muuttujan matemaattinen riippuvuus toisesta muuttujasta. Faktatietokantojen tietosisältö jäsentyy luonnollisella tavalla tiedoksi olioista, niiden ominaisuuksista ja suhteista. Näissä käsitteissä on olennaista, että ne eivät ole sidok-

sisä mihinkään perinteisistä tietomalleista: tietokannan käyttäjä voi näillä käsitteillä jäsentää tarkastelemaansa maailmaa ilman, että hänen tarvitsee samalla kiinnittää tietojen talletukseen huomiota.

On syytä erityisesti huomata, että relaatiomalliin (esim. [19] [22] [31] [34]) perustuvat ja etenkin sen normalisointiteoriaa soveltavat tietokannat sisältävät useita tietojen tallettamiseen ja käsittelyyn sidottuja ratkaisuja, jotka ovat selkeästi oliotarkastelutapaa teknisemmällä tasolla. Vaikka relaatiomallia mainostetaan hyvin tietoriippumattomaksi (so. käyttäjän ei tarvitse tuntea tiedon teknistä talletustapaa), eivät normalisoidut relaatiot läheskään aina ole käyttäjälle selkein ja luonnollisin tapa tietojen jäsentämiseen (esim. [36] [35]).

Käytettäessä faktahaun välittäjäjärjestelmää koko kysely »Anna niiden tuotteiden tavaramerkki, hinta ja myyntimaa, jotka ovat partavesiä, joita myydään Euroopassa ja joiden myyntiajankohta on vuosi 1989» pitäisi voida esittää käyttäjän jäsentämien olioiden perusteella (Kuva 4.1). Kuvan 4.1. kuvitteellinen kyselykieli muistuttaa SQL-kieltä (esim. [34]), mutta ei kohdistu relaatioihin vaan olioihin.

REPORT lotion(trademark, country-of-sale, price)
WHERE lotion **IS SOME** product
AND lotion(product-type) = 'after-shave-lotion'
AND lotion(year-of-sale) = 1989
AND lotion(country-of-sale) **IS PART OF** Europe

Kuva 4.1. Oliokyselyn mahdollinen esitystapa.

Kuvan 4.1. oliokyselykielen ilmaisu on rakenteeltaan seuraava. Avainsanan *REPORT* jälkeen esitetään tulostusmääritys ja avainsanan *WHERE* jälkeen ehtolauseke. Tulostusmäärityksessä määritellään tulokseen tulevien olioiden attribuutit. Tässä tapauksessa olioitu on vain yksi ja sen nimi on *lotion*. Siitä halutaan tulokseen attribuutit tavaramerkki, myyntimaa ja hinta, joista esimerkissä käytetään englanninkielisiä nimiä *trademark*, *country-of-sale* ja *price*.

Ehtolausekkeella esitetään ne ehdot, joiden vallitessa lotion-olioiden mainittujen ominaisuuksien arvot halutaan tulokseen. Ensinnäkin edellytetään, että lotion-oliot kuuluvat oliotyypin *product* (tuote). Oliotyyppimääritykseen käytetään avainsanaa *IS SOME*. Lisäehtoina edellytetään, että lotion-olioiden (1) tuotetyyppi

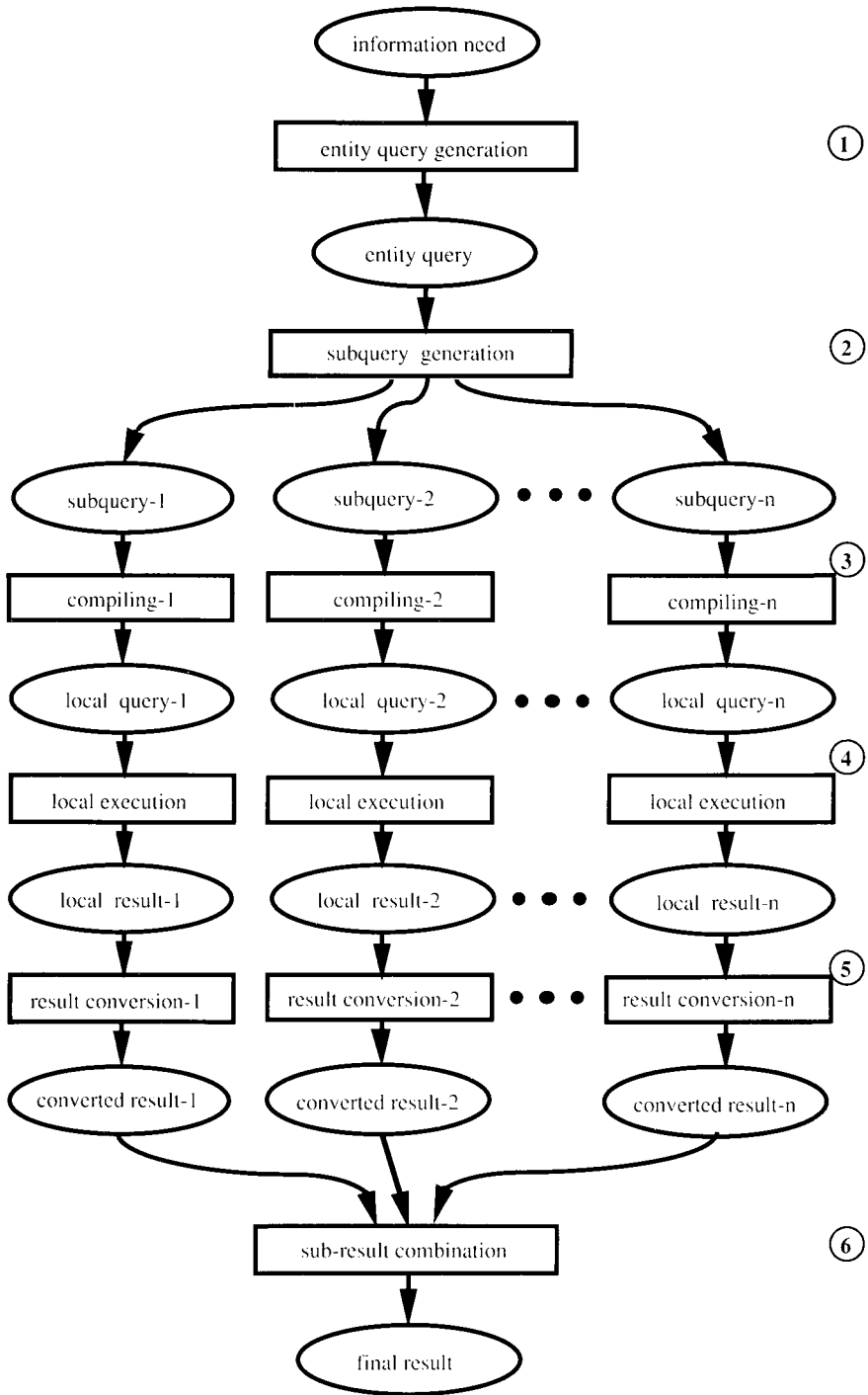
(ominaisuus *product-type*) on partavesi: *lotion*(*product-type*) = *after-shave-lotion*; (2) myyntiaika (ominaisuus *year-of-sale*) on vuosi 1989: *lotion*(*year-of-sale*) = 1989; (3) myyntimaa (ominaisuus *country-of-sale*) sijaitsee Euroopassa, minkä seikan ilmaisemiseen käytetään avainsanaa *IS PART OF*. Ehtolausekkeen osia yhdistää looginen konjunktio ('and').

4.2. Välittäjäjärjestelmän avustama hakuprosessi

Kyselyn esittämistä tulisi seurata kuvassa 4.2 kuvattu hajautettu tiedonhakuprosessi. Käsitelyssä voidaan erottaa kuusi päävaihetta: (1) tiedontarpeen täsmentäminen ja oliokyselyn luonti, (2) oliokyselyn osittaminen eri tietokantoihin, (3) tietokantakohtaisten osakyselyjen kääntäminen paikallisille kyselykielille, (4) paikallisten osakyselyjen suorittaminen, (5) paikallisten tulosten muuntaminen rakenteellisesti ja sisällöllisesti yhteensopiviin muotoihin, ja (6) paikallisten tulosten yhdistäminen. Tarkastelemme seuraavassa vaiheita lähemmin.

- *Vaihe 1.* Tiedontarpeen täsmentämiseksi käyttäjän tulee voida tutkia olioituja, niiden ominaisuuksia ja olioiden suhteita kuvaavia kiinnostavia käsitteitä oikeaa spesifisyydestä olevien käsitteiden tunnistamiseksi. Oliokyselyn luominen tulee edellyttää vain olioiden, niiden kiinnostavien ominaisuuksien ja olioiden suhteiden ilmaisemista.

- *Vaihe 2.* Kyselyn osittaminen eri tietokantoihin edellyttää tietokannan valintaa. Koska täsmällistä globaalia kuvausta tietokantojen sisällöstä ei ole olemassa, tulee välittäjäjärjestelmän kysyä lupaavista tietokannoista, sisältävätkö ne tietoa kyselyn olioituja, jne. Näin välittäjäjärjestelmä joutuu neuvottelemaan tietokantojen (tiedonhallintajärjestelmien) kanssa haun tai sen osien käsittelystä. Lupaavien tietokantojen tunnistaminen edellyttää tietokantojen korkean tason relevanssikuvauksista (aihepiiri ja muu kate) samaan tapaan kuin viitetietokantoja koskeissa hakemistoissa (niin painetuissa [14] kuin elektronisissakin). Vain näin voidaan välttää kyselyn levittäminen koko verkkoon (broadcasting) ja sen sijaan kohdentaa se todennäköisesti hedelmällisimpiin kohteisiin (narrowcasting) ja vain näin voidaan tukea osavastausten yhdistämistä monesta tietokannasta (esim. tuotteen hintatieto yhdestä paikasta ja myyntimäärätieto toisesta paikasta).



Kuva 4.2. Kyselyn käsittelyprosessi hajautetussa heterogeenisessä ympäristössä.

Tietokantojen vastaus olisi *yhteisellä kielellä esitetty kuvaus* oliokyselyssä mainittujen olioiden, ominaisuuksien ja suhteiden tai niiden alalajien katteesta tietokannassa. Näistä kuvauksista ilmenee olioita koskevien tietojen sisältö (mittayksiköt, -tarkkuudet ja ulottuvuudet, kate, jne.). Näistä tiedoista välittäjäjärjestelmä voi puoliautomaattisesti, yhdessä käyttäjän kanssa valita kyselyn kohdetietokannat. Tämän jälkeen kyselyn osittaminen kullekin kohdetietokannalle on mahdollista jo tunnetuilla menetelmillä.

- *Vaihe 3.* Tietokantakohtaisten olioihin perustuvien osakyselyjen kääntäminen kunkin paikallisen tiedonhallintajärjestelmän kyselykielelle edellyttää olioiden ja tietokannan sisällön yhdistävää kuvausmekanismia. Järvelin ja Niemi ovat tarkastelleet esimerkinomaisesti oliokyselyjen kääntämistä relaatiomallin kyselykielelle [18]. Sama periaate kuitenkin sopii vaikka tietokannan tietomalli olisi jokin kuin relaatiomalli.

- *Vaihe 4.* Tietokantakohtaisen, paikallisen kyselykielen kyselyn suorittaa paikallinen tiedonhallintajärjestelmä. Sen tuottama tulos on rakenteensa osalta sidoksissa tietokannan tietomalliin ja sisältönsä osalta paikallisiin tapoihin ja mittayksiköihin.

- *Vaihe 5.* Paikallisten tulosten muuntaminen rakenteellisesti ja sisällöllisesti yhteensopiviin muotoihin perustuu rakenteellisiin ja sisällöllisiin muunnossääntöihin, jotka perustuvat rakennepiirteiden formaaliin kuvaamiseen (esim. [24]) ja abstrakteihin tietotyyppisiin.

- *Vaihe 6.* Paikallisten tulosten yhdistäminen tarkoittaa niiden siirtoa laskentasolmuun tai -solmuihin, joissa alkuperäisen oliokyselyn edellyttämä laskenta voidaan suorittaa. Tietojen siirto voi tapahtua vaiheittain useamman solmun kautta, joissa välituloksia vaiheittain yhdistellään.

4.3. Välittäjäjärjestelmän tehtäväalueet

Hajautettujen faktatietokantojen käytön yksinkertaistaminen edellyttää, että faktanhaun välittäjäjärjestelmä hallitsee monia tehtäväalueita (ks. Taulukko 4.1, vrt. myös [20]). Järvelin ja Niemi ovat eritelleet näitä ominaisuuksia, niiden toteuttamisen ongelmia nykytiedon valossa sekä niihin liittyviä tietämystyyppisiä [18].

- Käyttöliittymän yksinkertaistaminen
- Keskustelukyky epämääräisen hakuaiheen tämentämiseksi
- Tietokannan valinta hajautettujen kuvausten avulla
- Kyselyn vuorovaikutteinen, puoliautomaattinen luominen käyttäjän hakuilmauksesta
- Huolenpito tietokannoista poimittujen tietojen yhteensopivuudesta
- Hakukustannusten ennakointi-, tarkkailu- ja kirjanpito
- Tietoliikenneyhteyden hallinta
- Tiedonhaun tekninen toteutus
- Haun tuloksen arviointi ja hakusuunnitelman parantelu
- Hakutuloksen jälkikäsitely ja tulkinta
- Käytön opastus ja käyttäjän ominaisuuksien, taitojen ja tapojen kuvaus (mallitus)

Taulukko 4.1. Faktahaun välittäjäjärjestelmän tehtäväalueet.

5. Tutkimusprojekti

Tutkimushanke, jonka tarkoitus on tuottaa kokeilu- ja tutkimusvälineistö faktahaun välittäjäjärjestelmiin liittyvien ongelmien tutkimiseen, on meneillään Tampereen yliopistossa tietojenkäsittelyopin ja kirjastotieteen ja informatiikan yhteistyönä. Sen tarkoitus ei ole tuottaa käytännössä toimivaa välittäjäjärjestelmää. Hanke on luonteeltaan konstrukttiivinen, metodologinen ja teoreettinen: huomio kohdistetaan ongelmien analysointiin ja niiden ratkaisemisessa käyttökelpoisten käsitteiden, jäsenysten, rakenteiden ja menetelmien luomiseen ja arviointiin abstrahoidussa ympäristössä faktahaun välittäjäjärjestelmää varten.

Tutkimusprojektissa huomio kohdistetaan lähemmin faktahaun välittäjäjärjestelmän seuraaviin tehtäväalueisiin:

- (1) Liittymän yksinkertaistaminen oliokeskeisen ja tietokantaympäristön tiedonhallintajärjestelmien tietomalleista riippumattoman käyttäjän tietomallin avulla.
- (2) Kyselyn luominen, olipa tietokannan tietomalli mikä tahansa tai käsittelytarve perinteinen tai rekursiivinen.
- (3) Kyky tukea epämääräisen hakuaiheen tämentämistä.
- (4) Tietokannan valinta hajautettujen kuvausten avulla.
- (5) Huolenpito tietokannoista poimittujen tietojen yhteensopivuudesta.

Huomion kohdistamista juuri näihin tehtäväläisiin tukee moni seikka. Ensinnäkin niihin liittyvät ongelmat ovat riittävän monipuolisia, vaikeita ja osin tuntemattomia, ettei niitä voida ratkaista suoraviivaisen kehitystyön avulla. Toiseksi nämä ongelmat ovat faktahaulle ominaisia ongelmia, joita ei esiinny esim. teksti- ja korviketietokantojen yhteydessä. Kolmanneksi, useita muita ominaisuuksia varten on käytökelpoista ja suoraviivaista kehitystyötä tukevaa tutkimusta jo olemassa esim. viitteenhaun välittäjäjärjestelmien yhteydessä (ks. [18]). Neljänneksi, luonnollisen kielen liittymä toisaalta ja tulosten tulkinta toisaalta ovat hyvin vaikeita ongelma-alueita, joiden tutkimisella (ratkaisemisesta puhumattakaan) ei ole kiirettä ennenkuin muut ongelmat on ratkaistu.

Faktahaun välittäjäjärjestelmien rakentamisessa ja toiminnassa tarvitaan monentyyppistä tietämystä ja sen esittämiseen sopivia esitystapoja. *Tietämys* (knowledge) on faktoja ja heuristiikkoja (asiantuntijoiden päättelysääntöjä). Ne voivat koskea mm. *olioita* ja niiden muodostamia luokkia, *tapahtumia* tai tietämystä itseään (*metatietämys*) [4] [16]. *Tietämyksen esitys* (representation) tarkoittaa sopivia tietorakenteita ja niitä tulkitsevia ohjelmia, jotka oikein käytettynä johtavat tietämystä ilmentävään käyttäytymiseen [4]. Tietämyksen *esitystapa* (myös representation) taas on joukko sopimuksia ja sääntöjä siitä, miten tarkasteltavia olioita tulisi kuvata [37].

Faktatietokantojen käytön yksinkertaistamiseen tähtäävä tutkimus on aidosti tieteiden välistä ja sijoittuu sekä kirjastotieteen ja informatiikan (erityisesti tiedonhaun tutkimuksen) että tietojenkäsittelyopin (erityisesti tiedonhallintatutkimuksen) yhteisalueelle. Esim. faktahaun välittäjäjärjestelmien rakentamisessa ovat käytökelpoisia ainakin seuraavat tiedonhallinta-, tekoäly- ja tiedonhaku tutkimuksen lähestymis- ja esitystavat sekä menetelmät:

- olioperusteiset tiedonhallintajärjestelmät ja kyselykielet
- tietokantojen tietomallien integrointi ja laajentaminen myös rekursiiviseen käsittelyyn
- semanttiset verkot ja kehysesitys
- abstraktit tietotyypit
- tiedonalojen tietämyksen käsitteellinen mallintaminen
- tietohakemistot

Olioperusteiset tiedonhallintajärjestelmät ja kyselykielet sekä tietokantojen tietomallien (kuten relaatio- ja hierarkkinen tietomalli) integrointi (esim. [5] [27] [29]) ja laajentaminen myös rekursiiviseen käsittelyyn (esim. [1] [2] [35]) tukevat välittäjäjärjestelmän liittymän yksinkertaistamisen ja tietomallien integroinnin tehtäväläisiin liittyvien ongelmien ratkaisemista. Semanttiset verkot ja kehysesitys (esim. [4] [37]) ovat tietämyksen esitystapaja, jotka sopivat kyselyn täsmennyksen tehtäväläisen ongelmien ratkaisemiseen. Niitä tarvitaan eri tiedonalojen tietämyksen käsitteelliseen mallintamiseen. Esimerkki tällaisesta mallintamishankkeesta on UMLS -projekti (Unified Medical Language System), jossa semanttisten verkkojen ja kehysesityksen avulla mallinnetaan lääketieteellistä tietämystä (esim. [7]). Semanttiset verkot ja kehysesitys voidaan liittää myös tietohakemistojärjestelmiin (esim. [8]) kuvaamaan tietokantojen tietosisältöä tietokannan valinnan ongelmien ratkaisemista varten. Yhdessä abstraktien tietotyyppien (esim. [9] [32]) kanssa semanttiset verkot ja kehysesitys tarjoavat esitystavat ja menetelmät tietojen yhteensopivuudesta huolenpidon ongelmien ratkointaan. Näitä lähestymis- ja esitystapoja sekä menetelmiä tarkastellaan yksityiskohtaisemmin faktahaun välittäjäjärjestelmien kannalta aikaisemmassa raportissamme [18].

Tietokantojen käyttäjien kannalta mainittuihin tehtäväläisiin kuuluvien tutkimusongelmien ratkaiseminen merkitsee sitä, että yhdessä muualla jo saavutettujen ratkaisujen kanssa saadaan *periaatteelliset valmiudet faktahaun välittäjäjärjestelmän toteuttamiseen*. Tästä on vielä pitkä ja työläs matka tietokantojen käyttäjiä *käytännössä* palvelevien faktahaun välittäjäjärjestelmien luomiseen. Periaateratkaisu ja sen toimivuuden testaaminen kuitenkin olennaisesti helpottavat tällaisen järjestelmän luomista.

Hyväksytty julkaistavaksi 3. 1. 1990.

Lähteet

- [1] Agrawal, R.; Jagadish, H.V., Direct Algorithms for Computing the Transitive Closure of Database Relations, Proceedings of VLDB Conference, Brighton, UK, 1987 : pp. 255—266.

- [2] Agrawal, R., Alpha: An Extension of Relational Algebra to Express a Class of Recursive Queries, Proceedings of the IEEE 3rd Data Engineering Conference, Los Angeles, 1987: pp. 580—590.
- [3] Auvinen, A. Kansainväliset markkinatietokannat viejän tietolähteinä. Helsinki, Finland: Suomen Ulkomaankauppaliitto, Tietopalveluryhmä; 1985.
- [4] Barr, Avron; Feigenbaum, Edward A., The Handbook of Artificial Intelligence, Volume I. London, UK, Pitman, 1981.
- [5] Bancilhon, F., Object-Oriented Database Systems. In: Proceedings of the 7th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, March 21—23, 1988, Austin, Tx : pp. 152—162.
- [6] Belkin, N.J. et al., Distributed expert-based information systems: an interdisciplinary approach. Information Processing and Management, 23(5): 395—409; 1987.
- [7] Barr, Charles E.; Komorowski, Henryk Jan; Pattison-Gordon, Edward; Greenes, Robert A., Conceptual Modelling for the Unified Medical Language System. In: Proceedings of the 12th Symposium on Computer Applications in Medical Care, Washington D.C., Nov. 1988. IEEE Computer Society, 1988.
- [8] Brathwaite, Ken S., Analysis, Design, & Implementation of Data Dictionaries. New York, NY, McGraw-Hill, 1988.
- [9] Breazeal, J.O.; Blattner, M.M.; Burton, H.D., Data standardization in heterogeneous data bases through the use of data abstraction. Information Processing and Management, 23 (5): 465—475; 1987.
- [10] Brooks, Helen M.; Daniels, P.J.; Belkin, Nicholas J., Research on the Information Interaction and Intelligent Information Provision Mechanisms. Journal of Information Science, 12(1—2): 37—44; 1986.
- [11] Chen, C.-c.; Herson, P. eds. Numeric Databases. Norwood, NJ, Ablex, 1984.
- [12] Croft, W.B., User-Specified Domain Knowledge for Document Retrieval. In: Rabitti, F. ed. Proceedings of the ACM Conference on Research and Development in Information Retrieval 1986, Pisa, Italy, Sept. 8—10, 1986: pp. 201—206.
- [13] Directory of Online Databases. Santa Monica, CA: Cuadra Associates; 1984.
- [14] Directory of Online Databases, 8 (1) 1987. New York, Cuadra/Elsevier, 1987.
- [15] Davies, R. ed. Intelligent Information Systems : Progress and Prospects. Chichester, UK, Ellis Horwood Ltd, 1986.
- [16] Feigenbaum, Edward A., Knowledge Engineering: The Applied Side of Artificial Intelligence. Stanford, CA, Stanford University, Department of Computer Science, Heuristic Programming Project, Report No. HPP-80-21, 1980.
- [17] Harter, Stephen P., Online Information Retrieval: Concepts, Principles, and Techniques. Orlando, FL, Academic Press, 1986.
- [18] Järvelin, Kalervo; Niemi, Timo, Hajautettujen faktatietokantojen käytön yksinkertaistaminen: ongelmia ja lähestymistapoja. Tampere, Finland, Tampereen yliopisto, Kirjastotieteen ja informatiikan laitoksen tutkimuksia, 32, 1989.
- [19] Järvelin, K., Numeeriset tietokannat ja niiden käyttökustannukset: uusia ongelmia informatiikalle ja informaatiopalvelulle. Kirjastotiede ja informatiikka, 5(2): 33—55; 1986.
- [20] Järvelin, Kalervo, A Blueprint of an intermediary System for Numeric Source Databases. IN: Koskiala, Sinikka; Launo, Ritva ed., Information* Knowledge* Evolution: Proceedings of the 44th FID Congress, Helsinki, Aug 28th — Sept 1st., 1988. FID Publication 675. Amsterdam, North-Holland, 1989: pp. 311—320.
- [21] Kahima, M.; Lukkari, M.; Mylly, H. Ulkomaisia suorakäyttöisiä tilastotietokantoja. Helsinki, Finland: Tilastokeskus, muistio N:0 98; 1985.
- [22] Maier, D., The Theory of Relational Databases. London, UK, Pitman, 1984.
- [23] Neufield, M. Lynne; Cornog, Martha, Database History: From Dinosaurs to Compact Discs. Journal of the American Society for Information Science, vol. 37(4): 183—190; 1986.
- [24] Niemi, T., Conversion of Flat Files and Hierarchical Data Bases. Acta Universitatis Tampereensis, ser A vol 194. Tampere, University of Tampere, 1985. (Ph.D. thesis).
- [25] Ojala, Marydee, Views on End-User Searching. Journal of the American Society for Information Science, 37(4): 197—203; 1986.
- [26] Pollit, A.S., CANSEARCH: An Expert Systems Approach to Document Retrieval. Information Processing and Management, 23(2): 119—138; 1987.
- [27] Roth, M.A., The Design of non 1NF Relational Databases Normal Form, Proceedings of ACM SIGMOD Conference, San Francisco, 1987; pp. 143—159.
- [28] Salton, G.; McGill, M.J. Introduction to Modern Information Retrieval. New York, NY, McGraw-Hill; 1983.
- [29] Scheck H.-J.; Pistor P., Data Structures for an Integrated Data Base Management and Information Retrieval System, Proceedings of VLDB Conference, Mexico City, 1982; pp. 197—206.
- [30] Seppälä, E.O., Tietohuolto tiedepoliitikassa. Korkeakoulutieto, 13(2): 6—7; 1986.
- [31] Soini, Tapani, Tietoanalyysi. Jyväskylä, Finland, Weilin + Göös, 1985.
- [32] Thomas, Pete; Robinson, Hugh; Emms, Judy,

- Abstract Data Types: Their Specification, Representation, and Use. Oxford, UK, Clarendon Press, 1988.
- [33] Tutkimuksen tietokannat -työryhmän muistio. Helsinki, Finland, Opetusministeriö, Työryhmien muistioita n:o 1986:22; 1986.
- [34] Ullman, J.D., Principles of Database and Knowledge Base Systems. Vol. I. Rockville, MD, Computer Science Press, 1988.
- [35] Ullman, J.D., Principles of Database and Knowledge Base Systems. Vol. II: The new technologies. Rockville, MD, Computer Science Press, 1989.
- [36] Wiederhold, G., Views, Objects, and Databases. *Computer*, 19(12): 37–44; 1986.
- [37] Winston, Patrick H., Artificial Intelligence. 2nd ed. Reading, MA, Addison-Wesley, 1984.