TEKNILLINEN KORKEAKOULU
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Tietoliikennetekniikan tutkinto-ohjelma

Antti Poikola

UUTISPALVELUIDEN TUOTTAMINEN HAKUTEKNOLOGIOIDEN AVULLA

Diplomityö

Valvoja: Akatemiaprofessori Mikko Sams
Ohjaaja: Elina Ruha

TEKNILLINEN KORKEAKOULU
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Tietoliikennetekniikan tutkinto-ohjelma

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Antti Poikola		
Työn nimi:	Uutispalveluiden tuottaminen hakuteknologioiden avulla		
Päiväys:	26. toukokuuta 2008	**Sivumäärä:**	6 + 69 + 2
Professuuri:	Kognitiivinen teknologia	**Koodi:**	S-114
Valvoja:	Akatemiaprofessori Mikko Sams		
Ohjaaja:	Elina Ruha		

Uutispalveluita on tieteellisesti tutkittu useimmiten vain loppukäyttäjän näkökulmasta. Uutistoimisto palveluiden tuottajana on mielenkiintoinen ja vähemmän tutkittu kohde. Tämän työn kohdeyrityksenä oli Suomen Tietotoimisto (STT). Johtoajatuksena tutkimukseen lähdettäessä oli, että STT:n journalistisesti tuottamasta materiaalista on mahdollista hakuteknologioita hyödyntämällä jalostaa uusia entistä paremmin käyttäjien tarpeita vastaavia uutispalveluita.

Tutkimus kuuluu informaatiotutkimuksen alaan. Informaatiotutkimuksessa on perinteisesti ollut toisistaan erillisinä tutkimusalueina tiedon loppukäyttäjään keskittyvä tiedonhankintatutkimus sekä tietokantoihin ja hakujärjestelmiin keskittyvä tiedonhakututkimus. Voimistuvana suuntauksena on yhdistää näitä kahta aluetta. Tässä työssä on mallinnettu Suomen Tietotoimistoa informaatiojärjestelmänä soveltaen yhdistettyä tiedonhankinnan ja tiedonhaun mallia.

Tutkimuksessa tehtiin katsaus nykyisiin hakuteknologioihin ohjelmistovertailun muodossa ja haastateltiin uutispalveluita käyttäviä asiakkaita. Näiden kahden osatutkimuksen pohjalta esitetään yleisellä tasolla ehdotuksia olemassa olevien uutispalveluiden kehittämiseksi. Ehdotuksissa on huomioitu haastattelututkimuksen yhteydessä selvitetyt loppukäyttäjien tarpeet ja sovellettu ohjelmistovertailun yhteydessä identifioituja hakuteknologioiden ominaisuuksia.

Yhdistetty tiedonhaun ja tiedonhankinnan malli soveltui hyvin jäsentämään uutistoimistoa monimutkaisena informaatiojärjestelmänä. Työn tuloksena syntyneet kehitysehdotukset ovat esimerkkejä siitä, miten hakuteknologioiden avulla voidaan tuottaa loppukäyttäjille hyödyllisiä uutispalveluita.

Avainsanat:	hakuteknologiat, uutispalvelut, uutistoimisto, yhdistetty tiedonhankinnan ja tiedonhaun malli
Kieli:	Suomi

HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation
Degree Programme for Communications Engineering

ABSTRACT OF MASTER'S THESIS

Author:	Antti Poikola		
Title of thesis:	Producing news services with search technologies		
Date:	May 26 2008	**Pages:**	6 + 69 + 2
Professorship:	Cognitive Technology	**Code:**	S-114
Supervisor:	Academy Professor Mikko Sams		
Instructor:	Elina Ruha		

News services have been scientifically studied mostly from the end users point of view. The news agency, as a producer of the news services, is an interesting research subject as well. This study is made for the Finnish national news agency STT. Leading idea behind the study is that it is possible to produce better news services from the journalistic material of STT by using the state-of-the-art search technologies.

This research belongs to the field of information science. Traditionally information science consists of two separate research areas: information seeking and information retrieval. Information seeking focuses on human as the end user of the information while information retrieval is concerned about the technical aspects of the information systems. Emerging trend is to combine these two research areas into one. In this study the Finnish national news agency has been modelled as an information system by using the integrated information seeking and –retrieval research framework.

First part of the study constitutes of the interviews made for the customers using news services in their work. Second part of the study is a comparison of the commercially available search technologies. Based on these two studies I propose general level improvements for the current news services. The proposals are based on the features of the modern search technologies which were found in search technology comparison. The end users needs that were found in the interview study are taken into consideration while making the proposals.

The integrated information seeking and –retrieval research framework suited well for the analysis of the news agency. The proposals of improvement are examples on how the search technology can be used to produce useful news services for the users.

Keywords: search technologies, news services, news agency, integrated information seeking and –retrieval research framework

Language: Finnish

Alkusanat

Tähän diplomityöhön johtanut tutkimus tehtiin vuoden 2006 aikana Suomen Tietotoimistossa.

Haluan kiittää työtäni valvoneita professoreita Iiro Jääskeläistä ja Mikko Samsia, joustavuudesta ja matkan varrella saamistani arvokkaista neuvoista diplomityön teossa. Haluan kiittää myös työtäni Suomen Tietotoimistolla ohjannutta päätoimittaja Atte Jääskeläistä, joka antoi minulle mahdollisuuden tutustua mediamaailmaan sisältä päin ja kannusti minua tutustumaan hakuteknologioita tarjoaviin yrityksiin juurta jaksaen. Kiitokset myös STT:n muulle henkilökunnalle, erityisesti Elina Ruhalle, Pauli Töllille ja Antti Pukerolle kaikesta saamastani tuesta. Työn loppuun saattamisessa minua ovat kannustaneet ja auttaneet erityisesti Tapio Takala, Inger Ekman, Jari Kätsyri ja Sari Karjalainen.

Espoon Otaniemessä 26.5.2008

Sisällys

1. Johdanto

Tutkimuksen tarkoituksena on tehdä katsaus siihen, minkälaisia käyttäjätarpeita on uutistoimiston palveluita käyttävillä median ulkopuolisilla toimijoilla, kuten yrityksillä, järjestöillä ja julkishallinnon organisaatioilla. Näissä organisaatioissa uutispalveluiden loppukäyttäjät ovat yleensä tiedotuksesta vastaavaa henkilökuntaa, hallintohenkilökuntaa tai tutkijoita.

Tutkimuksen kohdeyrityksenä on Suomen Tietotoimisto (myöhemmin STT). Tutkimus tarjoaa STT:lle käyttökelpoista siitä, minkälaisille hakuteknologioiden mahdollistamille uutispalveluille on eniten kysyntää uutistoimiston nykyisten ja tavoiteltujen asiakkaiden näkökulmasta. Työssä käsitellään STT:tä ja sen tuottamia palveluita informaatiojärjestelmänä. Palveluiden loppukäyttäjien näkökulma on tutkimuksessa keskeisellä sijalla. Tutkimus kuuluu informaatiotutkimuksen (information science) alaan.

Tutkimuksessa korostuu käyttäjien tilannesidonnainen informaatiokäyttäytyminen. Palveluita käyttävät asiakkaat ovat kognitiivisia toimijoita, joten kognitiotieteellinen näkökulma on työssä myös vahvasti esillä. Loppukäyttäjien ohella toisena tarkastelun kohteena on informaatiojärjestelmän tekninen parantaminen hakuteknologioiden avulla. Näin ollen tämä tutkimus sijoittuu kognitiotieteiden ja informaatiotekniikan rajamaastoon.

Tässä luvussa esitellään tutkimuksen lähtökohdat (Kappale 1.1) ja tutkimuksen rakenne (Kappale 1.2).

1.1. Tutkimuksen lähtökohdat ja taustat

STT:n media-asiakkaiden tarpeita ja tyytyväisyyttä on tutkittu säännöllisin välein STT:n media-asiakkaille suunnatuilla kyselytutkimuksilla. Tässä työssä aihetta lähestytään median ulkopuolisten käyttäjien eli pääasiassa eri organisaatioissa uutisointia seuraavien tiedottajien, analyytikoiden ja muun henkilökunnan näkökulmasta.

Tutkimuksen ulkopuolelle rajataan ne muutokset, joita tutkittujen palveluiden toteuttaminen vaatisi STT:n työprosesseissa. Myöskään taloudellisiin vaikutuksiin ei oteta kantaa, vaan tutkimus keskittyy nimenomaan loppukäyttäjien tarpeisiin ja teknisiin mahdollisuuksiin niiden tyydyttämiseksi.

1.1.1. Uutispalvelut

Uutistoimiston näkökulmasta toimituksellisesti tuotettu uutismateriaali on raaka-ainetta, jota pyritään kustannustehokkaasti jalostamaan mahdollisimman hyvin asiakkaita miellyttäviksi uutispalveluiksi. Uutispalveluilla tarkoitetaan samasta lähdeaineistosta, eli uutisjutuista, tiedotteista ja niihin liittyvistä kuva-, ääni- ja videodokumenteista valikoimalla tuotettuja kokonaisuuksia. Uutisia voidaan välittää eri formaateissa ja kohdentaa asiakasryhmien tarpeiden mukaisesti. Esimerkkeinä erilaisista uutispalveluista voidaan mainita mm. suoraan sanomalehtien toimitusjärjestelmään yhdistetty reaaliaikainen koko uutistoimiston tuotannon

kattava uutisvirta ja toisaalta yksittäiselle tilaajalle tekstiviestinä lähetettävät uutisotsikot hänen valitsemistaan aihealueista.

Uutispalveluiden kehittämisen lähtökohtana voidaan pitää sitä että, asiakas haluaa usein uutisia joltain rajatulta kiinnostuksen alueelta, mutta toisaalta hän haluaa kaikki häntä kiinnostavat uutiset helposti ja mahdollisimman nopeasti. Mitä paremmin uutispalvelu pystyy vastaamaan tähän haasteeseen, sitä suuremman arvon palvelu saa käyttäjien näkökulmasta. Tätä relevanttiuden haastetta kuvaa osuvasti alla oleva Atte Jääskeläisen (STT:n toimitusjohtaja vuonna 2006) kommentti.

"Toimittajat haluavat vain kiinnostavia tiedotteita, mutta he haluavat kaikki kiinnostavat tiedotteet." -Atte Jääskeläinen

1.1.2. Hakuteknologiat ja hakupalvelut

Tiedonhakuteknologioilla tarkoitetaan niitä ohjelmistoja ja algoritmeja, jotka mahdollistavat tehokkaiden hakupalveluiden tuottamisen johonkin laajaan lähdedokumenttien joukkoon. Hakuteknologioiden piiriin kuuluu mm. haettavien piirteiden, kuten avainsanojen erottaminen lähdemateriaalista, lähdedokumenttien muokkaaminen ja indeksointi, indeksin ylläpito, hakulausekkeiden muokkaus ja tulkinta, hakujen suorittaminen indeksiin, tulosdokumenttien järjestäminen jne. Hakuteknologiat kehittyvät kaiken aikaa yhä monimutkaisemmiksi ja laajenevat mm. kattamaan paljon suuremman kielivalikoiman; sen rinnalla kehittyvät kuvien, äänien ja liikkuvien kuvien hakuteknologiat. Samaan aikaan hakualgoritmien tehon kasvaessa ja sovellusalueiden laajentuessa myös niiden avulla tuotettujen palveluiden käytettävyys paranee.

Nykyiset tiedonhakuteknologiat, pääasiassa tekstihaut, ovat tulleet suurelle yleisölle tutuiksi internetin hakupalveluiden myötä. Käyttäjä onkin yleensä tekemisissä suoraan vain hakupalvelun kanssa, eikä hän edes tiedä, millä teknologioilla kyseinen palvelu on toteutettu.

Hakupalvelun käyttäjän kokema palvelunlaatu riippuu sekä sisällön laadusta että teknisestä laadusta. Sisällön laadulla tarkoitetaan sitä, mihin lähdeaineistoon palvelun kautta pääsee käsiksi ja kuinka laadukasta tietoa sieltä ylipäätään on mahdollista löytää. Tekninen laatu tarkoittaa sitä, kuinka hyvin palvelu on toteutettu ja pystyy auttamaan käyttäjää löytämään haluamansa. Samalla kun internetissä ilmaiset hakupalvelut ovat parantuneet entisestään, myös asiakkaiden vaatimustaso maksullisia palveluita kohtaan on kasvanut.

Uutistoimistossa etenkin median ulkopuolisilla asiakkailla korostuu tiedontarpeiden kapea-alaisuus, jolloin on oleellista, että kaikesta tuotetusta materiaalista pystytään tavalla tai toisella löytämään juuri ne uutiset, jotka asiakasta kiinnostavat. Hakuteknologioiden avulla on mahdollista monella tavalla auttaa käyttäjää saamaan haluamansa tieto esille ja näin ollen hakupalvelut ovat kiinteä osa myös uutispalveluita.

1.1.3. Käyttäjänäkökulma

Tämän tutkimuksen lähtöoletuksina on, että STT tuottaa nykyisellään paljon laadukasta sisältöä, joka kiinnostaa asiakkaita ja toisaalta, että hakuteknologioiden viimeaikainen kehitys on tuottanut varmasti paljon sellaista, mitä voitaisiin ottaa uutispalveluiden tuotannossa tehokkaaseen käyttöön. Kysymykseksi jää, miten STT voisi nykyisellä tavallaan tuottamasta lähdemateriaalista jalostaa paremmin

käyttäjien tarpeita vastaavia palveluita tämänhetkistä huipputasoa edustavien hakuteknologioiden avulla.

Tässä kysymyksenasettelussa uutispalveluiden käyttäjä on keskeisellä sijalla. Käyttäjän näkökulmasta palveluiden sisältö, toteutus, käytettävyys, teknologia ja joissain tapauksissa hinnoittelukin nivoutuvat tiiviisti yhteen. Toisaalta käyttäjän kokemukseen vaikuttavat myös monet muut tekijät, kuten hänen omat päämääränsä, aikaisemmat kokemuksensa, motivaationsa, tapansa, kulttuurilliset seikat, palvelun käyttötilanne jne.

1.2. Tutkimuksen rakenne

Luvuissa 2,3 ja 4 esitellään tutkimuksen kannalta keskeisimmät teoriat, terminologia ja aiheeseen liittyvää aikaisempaa tutkimusta. Ensin käydään läpi uutispalveluihin ja erityisesti uutistoimistoihin liittyviä tutkimuksia luvussa 2, sen jälkeen käsitellään hakuteknologioihin liittyvää käsitteistöä luvussa 3 ja teoriaosuuden lopuksi esitellään informaatiokäyttäytymisen ja informaatiojärjestelmien tutkimuksessa vallitsevia näkökulmia luvussa 4.

Luvuissa 5,6 ja 7 käsitellään tutkimusasetelma, tulokset ja vastaukset tutkimuskysymyksiin. Luvussa 5 esitellään tutkimusongelma sekä käydään läpi tutkimuksen kulku, menetelmät ja aineistot. Luvussa 6 esitetään Suomen Tietotoimiston mallinnus informaatiojärjestelmänä sekä raakatulokset kahteen osatutkimukseen (käyttäjähaastattelut ja ohjelmistovertailu). Luvussa 7 tehdään johtopäätöksiä saaduista osatutkimusten tuloksista nojautuen STT:n informaatiojärjestelmän malliin ja esitetään niiden pohjalta kolme konkreettista kehitysehdotusta. Lopuksi kappaleessa 8 vedetään lyhyesti yhteen koko tutkimuksen anti.

2. Informaatiotutkimus

Informaatiotutkimus tarkastelee tiedon välittymistä ihmisten, organisaatioiden ja yhteiskunnan toiminnassa ja käsittää niin ihmisten, organisaatioiden, kuin tekniikankin tutkimusta. Tieteenala on kehittynyt kirjastojärjestelmien tutkimuksesta laajemmin kaikkia informaatiojärjestelmiä koskevaksi tieteenalaksi. Informaatiotutkimuksen alueina voidaan erottaa tietohallinto (information management), tiedonhaku (information retrieval) ja tiedonhankinta (information seeking). Tietohallinto keskittyy organisaatiossa olevan informaation ja tiedon tarkoituksenmukaiseen hallintaan, mutta se ei ole tässä tutkimuksessa kiinnostuksen kohteena. Sen sijaan informaatiokäyttäytyminen (information behaviour) ja sen alakäsitteet; tiedonhaku ja tiedonhankinta ovat keskeisiä käsitteitä tässä tutkimuksessa.

2.1. Mitä on tieto?

Suomen kielessä termien data, informaatio (information) ja tieto (knowledge) arkikäytössä ei ole selvää eroa. Analysoitaessa informaatiojärjestelmiä ja tutkittaessa ihmisten informaatiokäyttäytymistä on kuitenkin olennaista tarkastella näiden termien eroja. Informaatiotutkimuksessa puhutaan yleisesti tiedon arvoketjusta (value chain of information), jossa irrallinen informaatio jalostuu ihmiselle käytännössä hyödylliseksi (kuva 1).

Kuva 1. Tiedon arvoketju (Haasio, Savolainen 2004)

Data on potentiaalista informaatiota, jollaista voi olla esimerkiksi asiayhteydestään irralliset faktat, joista informaation tuottajan toimesta voidaan jalostaa merkityksellisempää informaatiota, jota puolestaan voidaan välittää edelleen (Vakkari 1999).

Informaatio (Information) tulee latinankielisestä sanasta *informare* (muotoilu; muotoon paneminen). Chen ja Hernon (Chen, Hernon 1982) määrittelevät informaation tarkoittavan kaikkia niitä faktoja, ideoita, dataa ja fiktiivisiä hengentuotteita, joita informaation tuottaja on muotoillut omasta tietämyksestään ja jotka on kommunikoitu eteenpäin. Sen lisäksi, että informaatio on välitettävänä olevaa tietoa, se on myös jotakin, jonka pelkkä havaitseminen ei vielä merkitse tulkintaa tai ymmärtämistä (Haasio, Savolainen 2004).

Informaatiotutkimuksen näkökulmasta informaation on täytettävä kaksi ehtoa:

Informaatio on tarkoituksellisesti tehty muunnos informaation tuottajan käsityksistä eli kognitiivisista rakenteista.

Vastaanotettaessa informaatio myös muokkaa vastaanottajan tietämyksen tilaa (state of knowledge).

Mikäli vain ensimmäinen ehdoista täyttyy, puhutaan potentiaalisesta informaatiosta tai datasta. Mikäli vain jälkimmäinen ehdoista täyttyy, puhutaan aistihavainnoista tai

luonnossa esiintyvistä signaaleista, joita ei tulkita informaatioksi. (Ingwersen, Järvelin 2005)

Tieto (knowledge) syntyy kun informaation vastaanottaja tulkitsee informaation, minkä seurauksena se yhdistyy osaksi hänen tietorakennettaan ja muuttaa sitä. Tieto on siis jotain asiaa kuvaavan semanttisen tai pragmaattisen informaation tulkinta ja siihen liittyvä merkityksenanto. Kun tietoa välitetään, muuttuu se väistämättä informaatioksi (sanoiksi, kuviksi, taiteeksi, tieteeksi) ja altistuu informaation vastaanottajan omalle tulkinnalle. Se miten kukin tulkitsee saamansa informaation, riippuu mm. yksilön kulttuurillisesta ympäristöstä ja aiemmista kokemuksista ja jopa hetkellisestä tilanteesta. Kun vastaanotettu ja tulkittu informaatio johtaa myös toimintaan syntyy osaaminen tai taitotieto (know-how).

Tietämys tiedon synonyymina on ihmisellä tietyllä hetkellä oleva ymmärrys itsestään ja ympäröivästä maailmasta (Haasio, Savolainen 2004). Tähän hetkelliseen ymmärrykseen kuuluvat kognitiot, ajattelu, tunteet, tietoinen ja tiedostamaton muisti (hiljainen tieto). Tietämyksenä voidaan pitää yksilön omaksumien tietojen kokonaisvarastoa, joka vaikuttaa omalta osaltaan siihen, miten uutta informaatiota tulkitaan.

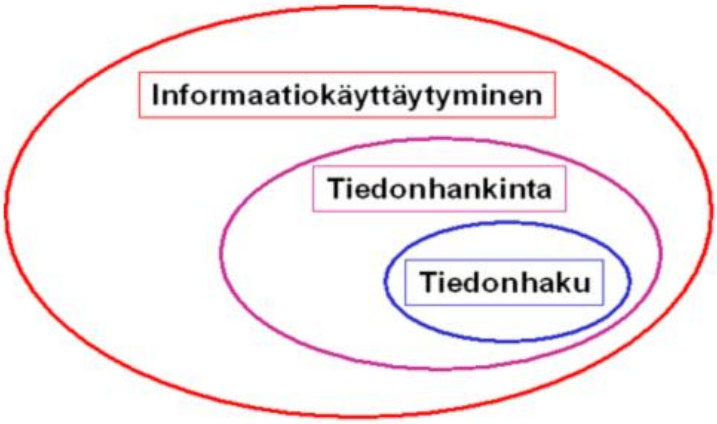
Viisaus (wisdom) on kyky hyödyntää tietämystä käytännön ongelmien ratkaisussa. Se on toiminnasta saadun kokemuksen ja osaamisen sekä tiedon ja ymmärryksen summa.

2.2. Tiedonhaku, tiedonhankinta ja informaatio-käyttäytyminen

Tiedonhakututkimus on lähempänä informaatiotekniikkaa ja keskittyy usein mikrotason ilmiöihin, esim. miten yksittäisestä hakuongelmasta (search task) muotoillaan hakulausekkeet (queries) ja miten löydettyjen dokumenttien relevanssia arvioidaan. Tiedonhankintatutkimus puolestaan on lähempänä ihmistieteitä ja suuntautuu voimakkaammin makrotason kysymyksiin, kuten millä perusteilla yksilöt valitsevat tietolähteitä ja hyödyntävät niitä eri tarkoituksiinsa.

Tiedonhaku suuntautuu tyypillisesti tietokantoihin, joihin syötetty tieto (linkit, uutiset, kuvat) jälleenhaetaan (retrieve) käyttöä varten. Tiedonhankinnassa on kyse laajemmasta läheisesti mm. oppimiseen ja ongelmanratkaisuun liittyvästä toiminnasta, jossa voidaan käyttää kaikkia mahdollisia tietolähteitä, kuten: henkilötietolähteitä, painettuja lähteitä ja tietokantoja yhdessä omaan kokemukseen perustuvan muistinvaraisen tiedon kanssa. Näin ollen tiedonhaku voidaan käsittää sinä osana tiedonhankintaa, joka voidaan toteuttaa tietokoneen avulla (kuva 2).

Voidaan ajatella, että tiedonhankinta käynnistyy jonkin tarpeen seurauksena ja päättyy lopulta löydetyn tiedon käyttöön. Tällaisia tiedonhankintaan johtavia motiiveja voivat olla esimerkiksi ajankohtaisten poliittisten tapahtumien seuraaminen, yksittäisen työtehtävään liittyvän ongelman ratkaiseminen tai uusien asioiden oppiminen (Savolainen 2000). Tiedontarpeisiin, -hankintaan ja -käyttöön voidaan viitata kokoavasti termillä informaatiokäyttäytyminen (Wilson 1997).



Kuva 2. Informaatiokäyttäytymisen hierarkkinen malli. (Wilson 1999)

2.3. Tiedontarpeet

Informaatiotutkimuksessa tiedontarpeet määritellään vastaamaan siihen, miksi tiedonhankinta käynnistyy ja mikä sitä ohjaa. Psykologiassa tarpeeksi määritellään epämiellyttävä olotila tai tuntemus, josta pyritään pääsemään eroon. Myös tiedonhankinnassa voidaan ajatella, että jotain asiaa koskeva ymmärryksen puute luo epävarmuutta, joka käynnistää tiedonhankinnan.

Tiedontarpeille on esitetty lukuisia erilaisia jäsennyksiä. Taylor (Taylor 1968) jäsentää tiedontarpeet jatkumona hyvin epämääräisistä ns. ydintarpeista, tietoisiin tarpeisiin, muotoiltuihin tarpeisiin ja lopulta ns. kompromissitarpeisiin. Belkin ja Dervin (Belkin 1984, Dervin 1993) vertaavat tiedontarpeita aukkoihin tai kuiluihin yksilön tietorakenteissa, jotka ilmenevät ongelmanratkaisun yhteydessä ja johtavat tiedonhankintaan aukkojen paikkaamiseksi tai siltojen rakentamiseksi. Tiedontarpeiden yksityiskohtainen määrittely on kuitenkin erittäin hankalaa ja useinmiten tiedontarpeilla viitataan kokoavasti kaikkiin niihin intresseihin, motiiveihin ja uskomuksiin, jotka käynnistävät ja ohjaavat tiedonhakua.

Eräs tapa tulkita tiedontarvetta on ajatella, että se pystytään identifioimaan yksityiskohtaisemmin vasta sen jälkeen, kun se on saatu tyydytettyä hankkimalla relevantiksi osoittautunutta tietoa (Savolainen 2000). Tämä näkökulma tuntuu varsin osuvalta esimerkiksi uutisseurannan tapauksessa. On tyypillistä, että tiedonhakijalla on jokin melko epämääräinen käsitys siitä, millainen uutinen saattaisi hänen kannaltaan olla hyödyllinen. Tämä epämääräinen käsitys ohjaa tiedonhakijaa, kun hän valitsee tietolähteitä, tekee hakuja, silmäilee uutisia ja etsii jotain mielenkiintoista. Vasta löydettyään mielenkiintoisen uutisen hän osaa sanoa, minkälaiseen tiedontarpeeseen se vastasi.

Myös jako orientoivan ja praktisen tiedontarpeen välillä soveltuu uutismaailmaan hyvin. Orientoivan tiedon tarpeet viittaavat pyrkimyksiin pysyä ajan tasalla ja seurata toimintaympäristön muutoksia kun taas praktisen tiedon tarpeet liittyvät jonkin ongelman ratkaisemiseen tai tehtävän menestykselliseen suorittamiseen (Savolainen 2000).

Uutiset sanan mukaisesti välittävät uutta informaatiota, joten yleensä uutisseurannalla pyritään vastaamaan orientoivan tiedon tarpeeseen. Voidaan todeta esimerkiksi, että "hänen on työnsä puolesta seurattava talousalan uutisointia", tällöin on kysymys orientoivan tiedon tarpeesta. Vanhat arkistoidut uutiset toisaalta

saattavat vastata myös johonkin praktisen tiedon tarpeeseen. Esimerkkinä praktisen tiedon tarpeesta voisi olla: "hänen täytyy selvittää paljonko LP levy keskimäärin maksoi nykyrahassa mitattuna vuonna 1976, jotta hän voisi kirjoittaa artikkelin äänitteiden hintakehityksestä musiikkialan lehteen".

2.4. Informaatiokäyttäytymisen ja -järjestelmien tutkimus

Informaatiokäyttäytymisen tutkimuksen alku voidaan jäljittää jo 1948 järjestettyyn *Royal Society Scientific Information Conference* konferenssiin, jossa esitettiin lukuisia aiheeseen liittyviä julkaisuja. Tämä tapahtui jo seitsemän vuotta ennen, kuin Chris Hanson muotoili termin "Information Science" (Wilson 1999). Tiedonhakututkimuksen alkuaikoina 1950- ja 60-luvuilla tehdyt ASTIA ja Cranfield tutkimukset loivat pohjan kokeelliselle tiedonhaun tutkimukselle (Cleverdon 1967, Ellis 1996). Myös ensimmäiset toimivat automaattiset tiedonhakujärjestelmät kehitettiin 60-luvun alkupuolella. Seuraavina vuosikymmeninä kehittyi kolme merkittävää lähestymistapaa tiedonhakututkimukseen: järjestelmäkeskeinen näkökulma, käyttäjäkeskeinen näkökulma ja kognitiivinen näkökulma.

Informaatiotutkimuksessa on aikakaudesta ja tutkijoista riippuen painotettu joko enemmän teknisiin järjestelmiin ja tiedonhakuun keskittynyttä tutkimusta tai sen vastapainona käyttäjään ja tiedonhankintaan keskittynyttä tutkimusta. Näistä ensimmäinen järjestelmäkeskeinen näkökulma on lähempänä tietotekniikkaa, kun taas käyttäjäkeskeinen tutkimus on ollut lähempänä kognitiotieteitä. Viimeaikoina on näitä kilpailevia näkökulmia pyritty myös yhdistämään ja tarkastelemaan informaatiojärjestelmiä kokonaisuuksina se. sekä tietotekniikka ja käyttäjät on huomioitu tasapuolisesti. Seuraavaksi käsitellään yksityiskohtaisemmin järjestelmä- ja käyttäjäksekeisenkeskeisien näkökulmien kehitystä sekä esitellään näitä yhdistävä tiedonhankinnan ja tiedonhaun malli.

2.4.1. Järjestelmäkeskeinen tekninen näkökulma

Järjestelmäkeskeisen tiedonhakututkimuksen kehitys on ollut teknologiavetoista. Tutkimuksen tavoitteena on ollut hyödyntää tietotekniikan kasvava potentiaali informaation prosessoinnissa kehittämällä aina parempia ja tehokkaampia hakualgoritmeja ja -järjestelmiä. Järjestelmäkeskeinen näkökulma perustuu ns. tiedonhaun laboratoriomalliin jonka keskeisinä osina ovat dokumentit, tietokannat, hakualgoritmit, hakulausekkeet ja tallennetut relevanssitiedot (relevance assesments). Järjestelmien tehokkuutta ja hyvyyttä on perinteisesti arvioitu hyvin kontrolloiduissa testiolosuhteissa ilman oikeita käyttäjiä. (Ingwersen, Järvelin 2005)

Viimeisen kahden vuosikymmenen aikana hakuteknologioiden teoreettinen kehitys samoin kuin käytännön sovellusten kehitys on ollut hämmästyttävän nopeaa. Text REtrieval Conference eli TREC-konferenssien aloittaminen 1990-luvun alkupuolella yhdessä tietokoneiden nopean kehityksen kanssa tarkoitti tiedonhakujen skaalaamista yhä suurempiin tietokantoihin ja kokotekstihakuihin pelkkien viitetietokantojen sijasta. Käytännön sovelluksissa Internet ja Internetin hakukoneet ovat mullistaneet tiedonhaun kentän sallimalla normaalikäyttäjille pääsyn valtavaan ja alati kasvavaan informaatiovarantoon. (Ingwersen, Järvelin 2005)

Laboratoriomallin vahvuus on ollut sen yksinkertaisuus ja yleinen hyväksyttävyys, joka on johtanut siihen, että tutkijat ovat voineet tukeutua toistensa työhön ja nopeasti kehittää uutta tietoa. Mallin heikkoutena pidetään sitä, ettei sen puitteissa voida sanoa mitään kehitettyjen algoritmien toimivuudesta tosielämän

käyttötilanteissa. Laboratoriomalli sulkee ulkopuolelleen mm. kaikki tiedontarpeisiin, tiedonhakutehtäviin ja itse tiedonhakijoihin liittyvät muuttujat. Malli ei tarjoa mitään teoreettista pohjaa selittämään, miksi joku hakuteknologia on menestyksekäs ja toinen ei jossain tietyssä tosielämän tilanteessa. (Ingwersen, Järvelin 2005)

Järjestelmäkeskeinen tiedonhakututkimus on tuottanut lukuisia tehokkaita algoritmeja, joilla voidaan suuristakin dokumenttikokoelmista löytää tiettyä Boolen hakulauseketta vastaavat dokumentit erittäin nopeasti. Käyttäjän vastuulle jää kuitenkin oikeanlaisen hakulausekkeen muotoileminen. Hakuteknologioiden kehitysvauhti on ollut niin suuri, ja niiden käytännön hyöty mm. Internetin hakupalveluissa on ollut niin kiistaton, että se on osaltaan haudannut alleen kysymyksen siitä, pitäisikö käyttäjää huomioida paremmin hakujärjestelmien arvioinnissa ja vertailussa.

Tiedonhakututkimus on laajentunut myös lukuisille uusille alueille mm.: automaattiseen tiivistelmien tekoon (text summarization), kysymyksiin vastaamiseen (question answering), tiedon suodattamiseen (filtering), monikieliseen hakuun (cross-language retrieval), aiheen havaitsemiseen ja seurantaan (topic detection and tracking), tekstinlouhintaan (text mining). Tekstin lisäksi myös puheen, musiikin, kuvien, videon ja hypermedian hakua on kehitetty aktiivisesti. Tämän tutkimuksen kannalta merkittäviä hakuteknologioiden kehityssuuntia käsitellään tarkemmin luvussa 3.

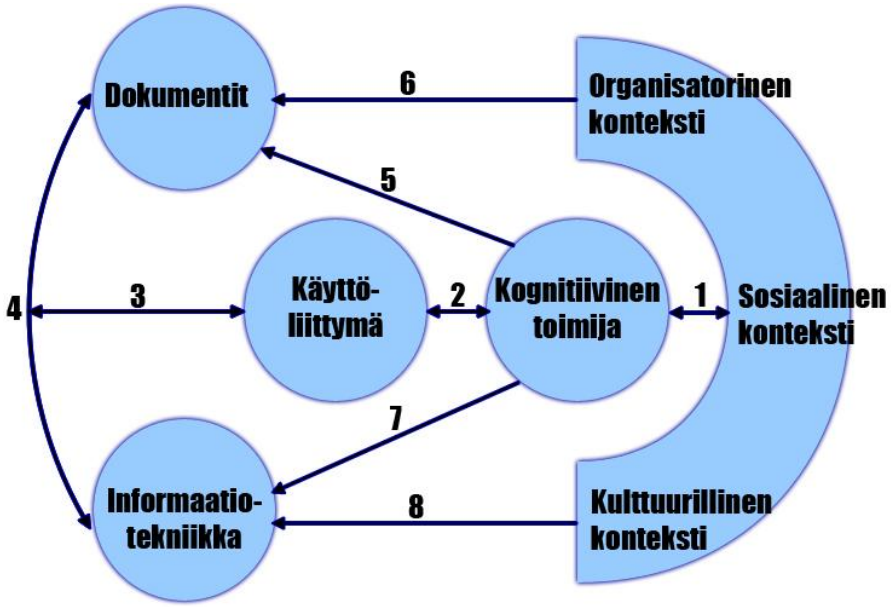
2.4.2. Käyttäjäkeskeinen kognitiivinen näkökulma

Brenda Dervin ja Michael Nilan (Dervin, Nilan 1986) nostivat artikkelissaan käyttäjän etusijalle järjestelmäkeskeisen informaatiotutkimuksen perinteestä poiketen. Dervinin ja Nilanin mukaan tieto ja informaatio ovat subjektiivisesti tulkittavia ilmiöitä, joiden merkitys vaihtelee tilanteittain. Uuden näkemyksen mukaan tiedon käyttäjä on aktiivinen toimija, joka etsii ja hyödyntää tietoa niistä lähteistä, jotka parhaiten tyydyttävät hänen tarpeitaan. Tiedonhankinta on osa ihmisten jokapäiväistä elämää ja jatkuvasti muutoksessa oleva prosessi, eikä irrallinen ilmiö. Käyttäjäkeskeinen näkökulma pyrkii huomioimaan tiedonhankinnan tilannesidonnaisuuden ja prosessiluonteen. Tärkeiksi nousevat ihmisten omat näkemykset siitä, miksi he valitsevat tiettyjä kanavia ja tietolähteitä ja miten he hahmottavat erilaisia tiedonhankinnan ongelmatilanteita. (Haasio, Savolainen 2004)

2.4.3. Yhdistetty tiedonhankinnan ja tiedonhaun malli

Tässä tutkimuksessa käytetään Peter Ingwersenin ja Kalervo Järvelinin (Ingwersen, Järvelin 2005) esittämää yhdistettyä tiedonhankinnan ja tiedonhaun tutkimuksen viitekehystä (Kuva 3). Viitekehys tarkastelee informaatiojärjestelmää kokonaisuutena, johon kuuluvat Informaatiosisältö, käytetyt IT ratkaisut, käyttöliittymä, sekä tiedon etsijät. Tässä tarkastelussa on merkittävää, että tiedon etsijät ovat kognitiivisia toimijoita, jotka toimivat sosiaalisen, kulttuurillisen ja organisaation ympäristön vaikutuksessa. Voidaankin puhua tiedonhausta kontekstissa. Tiedonhakijan toimintaan ja vuorovaikutukseen informaatiojärjestelmän muiden osien kanssa vaikuttaa mm: kuinka paljon hän tietää aiheesta, mikä on hänen tunnetilansa, mitkä ovat hänen tavoitteensa, miten hän itse ymmärtää sen, mitä hänen pitäisi tehdä jne.

Perinteistä tiedonhaun tutkimusta, jossa on vertailtu, mikä algoritmi palauttaa jollain mittarilla parhaiten hakulauseketta vastaavat dokumentit on siis laajennettu käyttäjän suuntaan ja toisaalta aiempaa käyttäjäkeskeistä tiedonhankinnantutkimusta on laajennettu huomioimaan paremmin myös IT ratkaisut. Järvelin ja Ingwersen kritisoivat informaatiotutkimuksen aikaisempia teoreettisia viitekehyksiä, jotka tarkastelivat ongelmia vain käyttäjän näkökulmasta (kognitive model) tai vain järjestelmien näkökulmasta (laboratory model) liian suppeiksi.



Kuva 3. Yleinen malli tiedonhankinnasta ja tiedonhausta, (Ingwersen, Järvelin 2005) . Mallissa keskeisellä sijalla ovat inhimilliset kognitiiviset toimijat, kuten esim. uutistoimittajat, hakualgoritmin suunnittelijat tai tiedon käyttäjät, jotka toimivat aina tila

Seuraavaksi selitetään, mitä tarkoitetaan informaatiojärjestelmällä ja kuvaillaan tarkemmin tiedonhankinnan ja tiedonhaun yleisen mallin eri komponentit: dokumentit, informaatiotekniikka, käyttöliittymä, kognitiivinen toimija ja konteksti.

Informaatiojärjestelmä

Informaatiojärjestelmällä tarkoitetaan tässä yhteydessä kokonaisuutta, jossa tietoa luodaan, tallennetaan ja siirretään loppukäyttäjille. Informaatiojärjestelmässä on osallisena useita kognitiivisia toimijoita, sekä tietoteknisiä komponentteja, jotka vaikuttavat toisiinsa joko suoraan tai ajan myötä. Ingwersenin ja Järvelinin mallissa (kuva 4) informaatiojärjestelmää voidaan tarkastella eri kognitiivisten toimijoiden näkökulmasta. Esimerkiksi toimittajan näkökulmasta: "toimittaja kirjoittaa uutisen, eli luo dokumentin" (nuoli 5) tai vastaavasti tiedon käyttäjän näkökulmasta: "käyttäjä hakee uutista hakusanalla käyttöliittymän kautta ja hakupalvelu palauttaa hakua vastaavan dokumentin, jonka käyttäjä lukee" (nuolet 2,3 ja 4).

Dokumentit

Dokumentit ovat sisällöllisiä kokonaisuuksia, joita informaatiojärjestelmän eri toimijat voivat tuottaa, muokata ja/tai etsiä. Dokumentteja voidaan ryhmitellä mm. niiden rakenteen, tyypin, tyylilajin, informaatiotyypin, viestinnällisen tarkoituksen, ajallisten piirteiden, merkkikielen, taiton ja tyylin, metadatan, sisällön tai linkitysrakenteen perusteella. (Ingwersen, Järvelin 2005)

On tärkeää huomata, että useinkaan yksittäinen dokumentti ei vastaa tiedontarpeeseen, vaan tietoa on haettava useista dokumenteista. Toisaalta yksittäinen dokumentti kokonaisuudessaan ei ole välttämättä hyödyllinen vaan ainoastaan joku pieni osa siitä saattaa olla tiedonhakijan näkökulmasta kiinnostava.

Ingwersen ja Järvelin käyttävät termiä informaatio-objekti vaihtoehtona sanalle dokumentti tarkoittaessaan yleensä digitaalisessa muodossa tallennettua kokonaisuutta, joka välittää potentiaalista informaatiota (Ingwersen, Järvelin 2005) . Tässä työssä käytetään termiä dokumentti sen selkokielisyyden vuoksi. Informaatiojärjestelmissä dokumentit muodostavat yhdessä niitä rikastavan indeksointi- ja metatiedon kanssa informaatioavaruuden, josta tietoa voidaan hakea.

Informaatiotekniikka

Informaatiotekniikalla tarkoitetaan kaikkia informaatiojärjestelmän laitteistoja ja ohjelmistoja, joidenka avulla tietoa voidaan hakea, säilyttää, indeksoida ja rikastaa automaattisesti.

IT-komponentin tärkeimmät tehtävät ovat:

- Dokumenttien fyysinen tallentaminen ja säilyttäminen sekä varmuuskopiointi
- Dokumenttien indeksointi, indeksin ylläpito ja hakujen suorittaminen
- Dokumenttien rikastaminen metadataa lisäämällä mm. luokittelemalla dokumentteja sisällön perusteella ja linkittämällä niitä toisiinsa

Käyttöliittymä

Käyttöliittymä on sen tärkeyden ja erityisen roolin takia esitetty mallissa informaatiotekniikasta erillisenä komponenttina. Pelkällä käyttöliittymällä ei toki voi tarjota käyttäjälle mitään sellaista tapaa päästä käsiksi informaatioon, mitä taustalla olevat muu informaatiotekniikka ei kykene toteuttamaan tehokkaasti. Yleisempi ongelma on kuitenkin, että käyttöliittymän heikkouksien takia käyttäjä ei todellisuudessa kykene hyödyntämään tarjolla olevaa informaatiota niin hyvin, mitä tekniikka periaatteessa mahdollistaisi.

Kognitiivinen toimija

Kognitiivisia toimijoita ovat kaikki informaatiojärjestelmässä osallisina olevat ihmiset. Kaikki kognitiiviset toimijat myös toimivat kontekstissa. Informaatiojärjestelmän kannalta merkittävimpiä kognitiivisia toimijoita ovat tiedon tuottajat, tiedonhakijat ja IT-komponenttien ja käyttöliittymien suunnitteluun osallistuvat henkilöt. Usein informaation valikointiin ja muokkaukseen ennen kuin se päätyy loppukäyttäjälle voi vaikuttaa tuottajan lisäksi muitakin kognitiivisia toimijoita kuten vaikkapa tiedon (manuaalinen) luokittelija, jonkin palvelun päätoimittaja, taittaja, informaatikko jne. Merkittävää on, että sama henkilö, sama kognitiivinen toimija voi tilanteesta riippuen olla eri roolissa esimerkiksi tiedon tuottajana tai tiedon luokittelijana.

Konteksti

Konteksti on hyvin yleisluonteinen käsite, jolla voidaan tieteellisessä kirjallisuudessa tarkoittaa melkein mitä tahansa. Tiedonvälityksen kontekstilla tarkoitetaan yleensä kokoavasti niitä varsinaisen informaatiosisällön ulkopuolisia seikkoja, jotka vaikuttavat välitettävän informaation tulkintaan ja merkityksen syntymiseen.

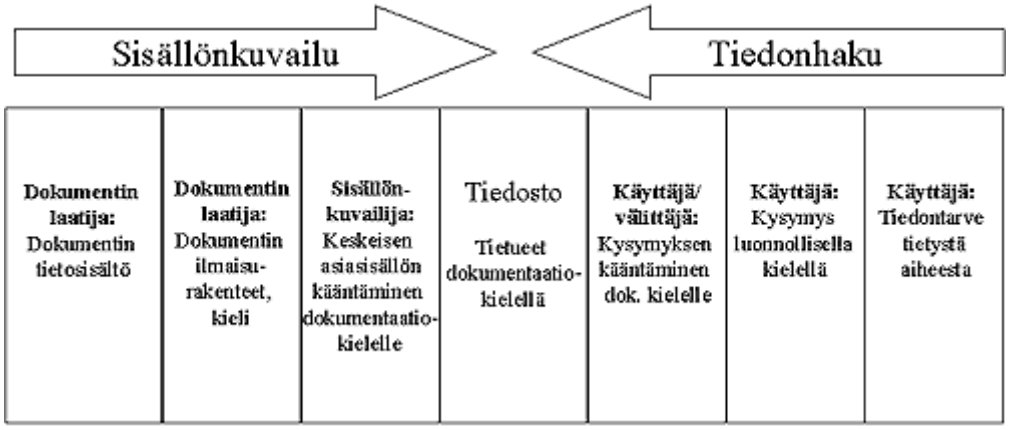
Kielitieteessä pragmatiikka tutkii tilanteen ja kontekstin vaikutusta merkityksen syntyyn. Merkitys käyttäjälle eli informaation tulkinta riippuu aina yhteydestä, jossa tietoa käytetään. Tulkintaan vaikuttaa siis dokumentin sisällön ohella lukuisat muut tekijät. Pelkkä tiedon tallennus ja siirto eivät riitä, jos tietoa ei pystytä tulkitsemaan tai se tulkitaan väärin (Bannon, Bødker 1997) . Tietoa on erittäin vaikea esittää sellaisessa muodossa, jossa kaikki tulkitsijat ymmärtäisivät tiedon tuottajan tarkoittamalla tavalla (Hertzum 1999).

Igversenin ja Järvelinin mallissa kontekstilla tarkoitetaan niitä organisatorisia, sosiaalisia ja kulttuurillisia seikkoja, jotka vaikuttavat suoraan tai ajan myötä tiedonhankintaan ja tiedonhakuun tutkittavassa informaatiojärjestelmässä.

3. Hakuteknologiat

Hakujärjestelmällä tarkoitetaan erityisesti tekstimuotoisen tiedon kuvailuun, tallennukseen ja hakemiseen suunniteltua tiedonhallintajärjestelmää (Alkula 2000) ja hakuteknologioilla dokumenttien automaattiseen käsittelyprosessiin liittyviä teknologioita, jotka yhdessä mahdollistavat vaivattoman informaation löytymisen käyttäjälle.

Perinteinen malli tiedonhausta esittää prosessin kaksi puoleisena, jossa toisella puolella on järjestelmä ja toisella käyttäjä (Kuva 4). Prosessin eri puolia nimitetään sisällönkuvailuksi (content description) ja tiedonhauksi (search). Sisällönkuvailun tuloksena hakujärjestelmässä on kokoelma dokumentteja, jotka on organisoitu ja esitetty tavalla, joka helpottaa niiden löytymistä. Tietoa haettaessa käyttäjillä puolestaan on tiedontarpeita, jotka he esittävät luonnollisella kielellä ja joista edelleen muodostetaan hakulausekkeita. Nämä kaksi puolta yhdistyvät pisteessä, missä hakulausekkeita verrataan organisoituihin dokumentteihin. Vertailun tuloksena saadaan lista hakulauseketta vastaavista dokumenteista, jotka esitetään käyttäjälle. Tätä perinteistä mallia on kritisoitu siitä, ettei se huomioi tiedonhaun interaktiivisuutta eli sitä, että lähes aina tiedonhaku on syklistä, käyttäjä tekee useita hakuja järjestelmästä ja kunkin haun tulokset muuttavat käyttäjän tiedontarpeita. (Robins 2000)(Alaterä, Halttunen & Sormunen)



Kuva 4. Tietokoneavusteinen informaationvälitysprosessi sisällöntuottajan ja tiedonhankkijan näkökulmista ja erivaiheisiin liittyviä hakuteknologioita. Mukailtuna (Salton 1989)

Yksi tiedonhaun tutkituimmista ongelmista on hakutulosten relevanssin määrittely. Relevanssilla tarkoitetaan sitä, kuinka hyvin dokumentti vastaa tiedonhakijan tiedontarpeeseen. Relevanssin käsitteestä voidaan edelleen johtaa haun hyvyyttä kuvaavat tarkkuuden ja saannin tunnusluvut. Tarkkuus on se osa hakutuloksista, jotka ovat relevantteja ja saanti on osuus kaikista relevanteista dokumenteista, jotka ovat hakutulosten joukossa. Relevanttiuden määrittely on hakujärjestelmille kaikkea muuta kuin triviaali ongelma, sillä käyttäjälle itselleenkin on usein hankala tehtävä tarkasti kuvailla omaa tiedontarvettaan. (Järvelin, Sormunen 1999)

Tässä luvussa esitellään tiivistetysti dokumenttien käsittelyprosessin eri vaiheet ja niihin liittyviä teknologioita. Teknologiat on jaoteltu tiedonhaun perinteisen mallin mukaisesti, kappaleessa 3.1 esitetään mallin keskellä oleva tietokanta, kappaleessa

3.2 mallin vasen puoli eli sisällön kuvailu ja kappaleessa 3.3 mallin oikea puoli eli tiedon haku. Soveltuvissa kohdin on huomioitu myös käyttäjän ja järjestelmän välinen vuorovaikutus, jota malli ei erityisesti tuo esiin.

3.1. Tietokanta

Tietokonepohjaisissa hakujärjestelmissä keskeisellä sijalla sisällönkuvailun ja tiedonhaun välissä on tietokanta (database), jonne dokumentit ja niihin liittyvät kuvailutiedot tallennetaan hakuja varten. Tietokanta koostuu joukosta tietueita (record). Tietueeseen on koottu yhdeksi käsiteltäväksi yksiköksi kaikki haettavaa kohdetta koskevat tiedot. Tietue koostuu kentistä (field), jotka voivat olla vapaamuotoisia tekstiä sisältäviä kenttiä, määrämittaista ja -muotoista kuvailutietoa sisältäviä kenttiä tai viittauksia muualle tallennettuun tietoon. (Alkula 2000)

Uutistietokannat on yleensä toteutettu niin sanottuina kokotekstitietokantoina, joissa koko uutisteksti on tallennettuna tietueeseen. Erotuksena kokotekstitietokantoihin ovat mm. kirjastojen käyttämät viitetietokannat, joissa haut voidaan kohdistaa vain tietokannassa oleviin julkaisun yksilöintitietoihin (nimi, tekijä, julkaisija) ja sisältöä kuvaaviin tiivistelmiin tai asiasanalistoihin, mutta itse julkaisut on talletettuna tietokannan ulkopuolelle eli kirjaston tapauksessa hyllyihin. Viitetietokantoja käytetään myös ei-tekstimuotoisen elektronisen tiedon hakuun, kuten kuva- ääni ja videotallenteiden hakuun. Tällöin tietokannassa on esimerkiksi äänitteen yksilöintitiedot tekstimuodossa ja linkki äänitiedostoon.

Tekstitietokannoissa tietueita voidaan nimittää myös dokumenteiksi ja Internetin myötä on dokumenttien kuvailutietojen nimityksenä yleistynyt metadata-termi. (Alaterä, Halttunen & Sormunen)

Tietokannan kenties tärkein ominaisuus on se, kuinka nopeasti sinne tallennettu tieto on mahdollista löytää. Kaikkein yksinkertaisin tapa löytää tietokannasta jokin tietty kohde on käydä yksitellen läpi tietueita, kunnes haluttu kohde löytyy. Tällä tavalla toimiva naiivi hakualgoritmi joutuu keskimäärin tarkastamaan puolet kaikista tietokantaan tallennetuista tietueista ja pahimmillaan kaikki ennen kuin se löytää mitään.

Tietokantojen toimintaa voidaan kuitenkin nopeuttaa merkittävästi indeksoinnin ja indeksiä eli hakemistoa hyödyntävien nopeampien hakualgoritmien avulla. Indeksillä tarkoitetaan yleisesti mitä tahansa tietorakennetta, jonka tarkoituksena on nopeuttaa hakuja. Vertauskuvana voitaisiin pitää tavallista arkistomappia, joka on tietokanta ja mapin värikkäitä välilehtiä, jotka nopeuttavat oikean sivun löytymistä ja toimivat siten indeksin lailla. Tavallisin tietokannoissa käytetty indeksi on tietueiden jonkun kentän mukaan järjestetty lista, josta löytyy osoittimet itse tietueisiin. Tällaisen indeksin avulla on mahdollista suunnitella hakualgoritmeja, jotka löytävät erittäin nopeasti tietyt ehdot täyttävät tietueet tietokannasta.

Tietokantojen koon kasvu ja nopeusvaatimukset ovat johtaneet siihen, että entistä tehokkaampia indeksointimenetelmiä ja hakualgoritmeja kehitetään jatkuvasti. Arvioitaessa vaihtoehtoisia indeksointitapoja pitää yleensä tehdä valintoja indeksin vaatiman koon, hakujen nopeuden ja indeksin päivitysnopeuden suhteen.

Kokoteksti-indeksoinnissa periaatteena on, että kaikki dokumentissa esiintyvät sanat tallennetaan hakemistoon. On kuitenkin syytä huomioida, että tiedonhakututkimuksen valtavirta suuntautuu englanninkielisen tekstin tulkitsemiseen (Alkula 2000). Kokoteksti-indeksointi kaikkein yksinkertaisimmillaan

ei sovellu morfologisesti monimutkaiseen suomen kieleen läheskään yhtä hyvin kuin englantiin, sillä kaikista lukuisista eri taivutusmuodoista muodostuisi omia hakusanoja hakemistoon ja toisaalta yksittäisten sanojen esiintymistiheys jäisi vastaavasti pieneksi.(Järvelin 1995)

3.2. Sisällönkuvailu

Kuvailumenetelmät ovat määriteltyjä käytäntöjä, joita sovelletaan liitettäessä yksittäisiä dokumentteja osaksi kokoelmaa. Dokumenttiin liitetyt kuvailut ovat tiedon organisoinnin konkreettinen perusta nykyaikaisessa tietokoneella hallitussa dokumenttikokoelmassa. Dokumenttien kuvailutiedot koostuvat yleensä sekä luettelointitiedoista että sisällönkuvailutiedoista. Luettelointitietoja ovat dokumentin ulkoisia piirteitä ja alkuperää kuvaavat tiedot kuten kirjoittaja, julkaisuajankohta ja –paikka, kun taas sisällönkuvailutiedoilla tarkoitetaan nimenomaan dokumentin sisältöä kuvaavia tietoja. Tiedonhakijan kannalta nämä molemmat palvelevat samoja päämääriä, tiedon löytämistä ja valikointia. Tässä paneudutaan erityisesti sisällönkuvailuun, jonka menetelmien piiriin kuuluvat mm. luokitusjärjestelmät, asiasanastot, linkitys, tiivistelmien teko jne. (Taulukko 1).

Taulukko 1. Sisällönkuvailun menetelmiä (Alaterä, Halttunen & Sormunen)

	Manuaaliset menetelmät	**Automaattiset menetelmät**
Dokumentaatiokieli (kontrolloitu sanasto)	asiasanoitus ja luokitus	automaattinen asiasanoitus ja luokitus
Luonnollinen kieli	avainsanoitus, tiivistelmät	kokoteksti-indeksointi, klusterointi, automaattiset tiivistelmät
Dokumenttien väliset suhteet	linkitys	automaattinen linkitys, viittausindeksointi, linkki-indeksointi

Suurin osa sisällönkuvailun menetelmistä on perinteisesti toteutettu manuaalisesti tiedon tuottajan tai jonkun toisen tiedon organisoijan toimesta. Sittemmin menetelmiä on kokonaan tai osittain automatisoitu. Informaatiotutkimuksen piirissä sekä manuaalinen että automaattinen sisällönkuvailu jaetaan yleisesti kahteen päälohkoon: luokitukseen ja indeksointiin.

Luokittelulla tarkoitetaan dokumenttien ryhmittelyä niiden sisältöä vastaaviin luokkiin ja se on tyypillisesti hierarkkista ja usein koodeihin perustuvaa. Luokituskoodeja ovat mm. kirjastojen käyttämä Yleinen Kymmenluokitus (UDK-luokitus) ja Deweyn luokitus. Luokittelu perustuu kuvailua varten kehitettyyn dokumentaatiokieleen eli luokitusjärjestelmään.(Järvelin 1995)

Indeksoinnilla tarkoitetaan prosessia, jossa yksi tai useampia asia- tai avainsanoja liitetään kuhunkin dokumenttiin (Belew 2000). Asiasanoittamisesta puhutaan, mikäli indeksointi tehdään dokumentaatiokielellä asiasanastoon tai tesaurukseen perustuen.

Indeksointia voidaan tehdä myös luonnollisella kielellä, jolloin puhutaan avainsanoituksesta. Tyypillistä on, että sekä asia että avainsanat muistuttavat luonnollista kieltä. Luokitus- ja indeksointijärjestelmillä on paljon yhteisiä piirteitä, eikä rajanveto niiden välillä ole tämän tutkimuksen kannalta oleellista.

Käyttäjän kannalta merkittävämpi ero on luonnolliseen kieleen perustuvien ja dokumentaatiokieleen (kontrolloitu sanasto) perustuvien kuvailumenetelmien välillä. Yleensä luonnollisella kielellä suoritettu sisällönkuvailu on käyttäjän kannalta helpompi ymmärtää ja se lisää hakijan mahdollisuuksia arvioida dokumenttia. Toisaalta luonnollisella kielellä tehdyt kuvailut ovat automatisoidun tiedonhaun kannalta ongelmallisia (Järvelin 1995) .

3.2.1. Sisällönkuvailu luonnollisella kielellä

Luonnollisella kielellä suoritettava sisällönkuvailu voi tapahtua sisältöä edustavilla avainsanoilla, poiminnolla, tiivistelmällä tai klusteroinnilla. Avainsana (keyword) on dokumentin sisältöä kuvaava, merkityksellinen sana tai termi, joka on poimittu dokumentin tekstistä joko automaattisesti tai manuaalisesti. Tiivistelmä on yleensä manuaalisesti tuotettu lyhyt esitys dokumentin sisällöstä. Nykyteknologialla varsinaisten tiivistelmien tuottaminen dokumenteista automaattisesti on hankalaa, mutta esimerkiksi Internetin hakupalvelut esittävät hakutulosten yhteydessä hakutermiä vastaavia automaattisesti tuotettuja poimintoja (extract) dokumenteista, jotka auttavat tiedonhakijaa valitsemaan tulosjoukosta itselleen hyödylliset dokumentit. Klusterointi eli ryvästäminen on automaattinen luokitusmenetelmä, jolla kootaan toisiaan muistuttavat dokumentit yhteen ryppäiksi. Klusteroinnissa ei käytetä ennalta määrättyä luokitusjärjestelmää, vaan luokittelu syntyy dokumenttijoukon sisällöstä ja perustuu näin ollen luonnolliseen kieleen (Järvelin 1995) .

3.2.2. Sisällönkuvailu dokumentaatiokielillä

Dokumentaatiokielillä tarkoitetaan kontrolloituja, yleensä jonkun ryhmän tai instituution kehittämiä tiedon kuvailuun tarkoitettuja sanastoja. Kontrolloidun sanaston tavoitteena on luoda tiedon kuvailijoille ja hakijoille yhteinen mahdollisimman yksiselitteinen kieli, jotta tallennuksessa ja haussa käytettävät ilmaisut kohtaisivat paremmin ja luonnollisen kielen monimuotoisuuteen liittyviltä ongelmilta vältyttäisiin.

Yksinkertaisimmillaan dokumentaatiokieli on aakkosellinen asiasanasto, josta asiasanoituksen yhteydessä valitaan kutakin dokumenttia parhaiten vastaavat asiasanat (controlled term, descriptor). Thesaruksella tarkoitetaan hierarkkista asiasanalistaa, jossa kuvataan myös termien välisiä suhteita, jolloin termeihin merkityksen perusteella liittyvät toiset termit voidaan löytää helpommin kuin esimerkiksi aakkosellisesta hakemistosta. Arkikielessä thesarus ja asiasanasto ovat nykyään synonyymejä. Esimerkki Suomalaisesta laajasta Thesaruksesta on YSA (Yleinen suomalainen asiasanasto)(Kansalliskirjasto). Luokittelussa käytettävät dokumentaatiokielet eli luokitukset eroavat asiasanastoista siinä, että niissä ei välttämättä pyritäkään edes luonnollista kieltä muistuttavaan esitykseen, vaan tärkeää on termien täsmällinen systemaattinen ja hierarkkinen esitys, jossa kullakin luokalla on yksiselitteinen symboli. Luokituksen käyttämiseksi tarvitaan usein aakkosellinen hakemisto, jossa aihepiiriä kuvaavalla sanalla voi etsiä sitä kuvaavan luokkasymbolin. Käyttäjän kannalta asiasanastot ovat usein helpompia, kuin luokitukset, mutta toisaalta luokitukset täsmällisesti määriteltyinä palvelevat

automaattista tietojen käsittelyä. Automaattisen tietojenkäsittelyn tarpeita varten luokituksia ja thesaruksia voidaan edelleen täsmentää ontologioiksi, joissa kaikki termien väliset suhteet on kuvattu koneen ymmärtämässä muodossa (Hyvönen 2005) .

Kontrolloidun sanaston käytöllä voidaan parantaa ja yhdenmukaistaa tiedon indeksointia, jolloin tiedon haussa päästään myöhemmin parempaa tarkkuuteen ja saantiin. Haittana on sanaston kehityksestä ja ylläpidosta aiheutuvat kustannukset sekä sanastotyön hitaus. (Hyvönen 2005)

3.2.3. Linkit sisällönkuvailuna

Kielellisten keinojen lisäksi myös dokumenttien välisten assosiatiivisten suhteiden esittämistä voidaan pitää sisällönkuvailuna. Manuaalinen tai automaattinen linkitys muihin dokumentteihin sekä jo olemassa olevien viittausten ja dokumenttiin johtavien linkkien indeksointi hyödyntää dokumenttien välisiä suhteita dokumentin kuvailuun.

Erityisesti verkkouutispalveluissa on tavanomaista esittää linkit samankaltaisiin tai samaa aihetta käsitteleviin uutisiin.

Kuva 5. Esimerkki bbc.com uutispalvelusta, joka hyödyntää linkkejä uutisen sisällön kuvailussa. Kuvassa oikeassa laidassa on "See also" linkkejä muihin aiheesta julkaistuihin uutisiin.

3.3. Tiedonhaku

Tarkasteltaessa perinteistä tiedonhakuprosessia (Kuva 4) tiedonhakijan näkökulmasta keskeinen kysymys on, miten tiedonhakija voi parhaiten esittää tiedontarpeensa hakujärjestelmän ymmärtämällä tavalla? Toinen tärkeä kysymys on, kuinka hakujärjestelmän löytämät tulokset kannattaa esittää tiedonhakijalle? Tiedontarpeiden ja hakutulosten esitystä voidaan ajatella myös käyttäjän ja järjestelmän välisenä vuorovaikutuksena, jossa käyttäjä antaa palautetta järjestelmälle ja järjestelmä käyttäjälle.

Tavallisesti tiedonhakija syöttää hakukenttään joitain avainsanoja tai tekstiä luonnollisella kielellä ja järjestelmä muokkaa tästä ns. hakulausekkeen, joka on

alkuperäisen tiedontarpeen representaatio sellaisessa muodossa, jota hakujärjestelmä voi tehokkaasti käyttää. Hakulausekkeet ovat aina epätäydellisiä, koska tiedontarpeiden määrittäminen on ihmisille luonnostaankin vaikeaa (Belkin, Croft 1987) ja hakujärjestelmän ymmärtämä dokumentaatiokieli on aina ilmaisuvoimaltaan rajoittuneempaa, kuin luonnollinen kieli. Tässä ns. aktiivisessa haussa perusajatuksena on, että tietokanta on suhteellisen staattinen tai hitaasti muuttuva, mutta tiedonhakijoiden tiedontarpeet muuttuvat kerrasta toiseen. Toinen lähestymistapa, joka sopii hyvin uutisaineistoon ja muuhun jatkuvasti päivittyvään tietovirtaan on profilointi. Profiloinnissa lähtökohtana on, että tiedonhakijan kiinnostuksenkohteista voidaan tehdä suhteellisen harvoin muutettava hakulauseke, johon kaikkea jatkuvasti tietokantaan tulevaa uutta aineistoa voidaan verrata automaattisesti.

Hakujärjestelmän ydintekniikan osalta aktiivinen haku ja profilointi eivät eroa merkittävästi toisistaan. Molemmissa tapauksissa dokumenteista tehdään representaatiot, joita verrataan hakulausekkeisiin ja tuloksena järjestelmä palauttaa ne dokumentit, jotka parhaiten vastaavat hakulauseketta. Myös samoja hakulausekkeiden muokkaukseen ja hakutulosten esittämiseen soveltuvia teknologioita voi käyttää molemmissa tapauksissa.

Taulukko 2. Tiedonhakuun liittyviä teknologioita

Aktiivinen haku	avainsanahaku, käsitehaku
Passiivinen haku (profilointi)	suodattaminen, reitittäminen
Vuorovaikutteinen haku	implisiittinen käyttäjän mallinnus, relevanssipalaute, drill-down, samankaltaisten haku
Haun muokkaus	termin laajennus, stop-sanojen poisto, oikoluku, hakulausekkeiden ja kohdedokumenttien kääntäminen (kieltenvälinen haku)
Tulosten esittäminen	tulosten järjestäminen, klusterointi, poiminnot, visualisointi

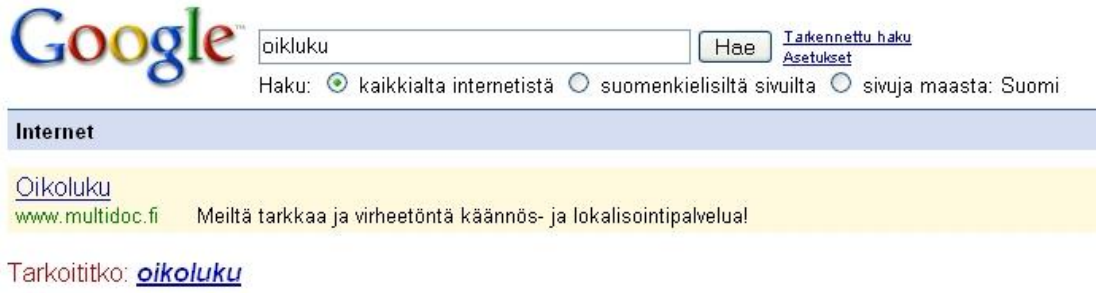
3.3.1. Aktiivinen haku ja haun muokkaus

Tekstitietokannoissa yleinen kokoteksti-indeksointiin perustuva vapaatekstihaku (vrt. esim. Google-haut) on erittäin tehokkaaksi kehittynyt ja tarjoaa hakumahdollisuuksien perustason minimikustannuksilla (Alaterä, Halttunen & Sormunen) .

Tavallisin lähestymistapa tiedonhakuun on avainsanahaku (keyword search), jossa käyttäjä syöttää tekstikenttään avainsanoja, joita hän olettaa löytyvän itseään kiinnostavista dokumenteista. Hakukone palauttaa sitten listan niistä dokumenteista, joista käyttäjän esittämiä hakusanoja löytyy. Avainsanahakua voidaan laajentaa ns. Boolen haulla, jolloin yksittäisistä avainsanoista voidaan yhdistää AND, OR ja NOT operaattoreilla monimutkaisempia hakulausekkeita (Heaps 1978).

Avainsanahaun tarkkuus paranee merkittävästi käytettäessä useampia hakusanoja ja monimutkaisempia hakulausekkeita. Tämä johtuu siitä, että yksittäinen hakusana esiintyy usein myös dokumenteissa, jotka eivät ole käyttäjän kiinnostuksen kohteena. Perusongelmana avainsanahaussa on, että haun onnistuminen riippuu käyttäjän taidosta ja viitseliäisyydestä tehdä hyviä hakulausekkeita.

Haun muokkauksessa hakukone pyrkii eritavoin parantamaan käyttäjän syöttämää hakulauseketta ennen tietokantahaun suorittamista. Yleisiä haun muokkauksia ovat mm. stop-sanojen poisto, termin laajennus ja oikoluku. Stop-sanojen poistossa hyvin yleiset ja sisällöllisesti merkityksettömät sanat, kuten: ja, ei, mutta, jne. poistetaan hakulausekkeesta. Termin laajennuksessa alkuperäisen hakusanan lisäksi haetaan myös sen eri taivutusmuotoja ja synonyymejä. Oikoluku puolestaan pyrkii tarkastamaan alkuperäisen hakulausekkeen oikeinkirjoitusta ja ehdottaa oikeinkirjoitettuja avainsanoja, mikäli vaikuttaa, että käyttäjä on tehnyt kirjoitusvirheitä. Osa muokkauksesta voi olla käyttäjälle täysin huomaamatonta, kuten stop-sanojen poisto ja osa interaktiivista, kuten oikolukuehdotukset, joita suosittu Google hakukonekin ehdottaa käyttäjälle.



Kuva 6. Esimerkki Google hakukoneen automaattisesti antamasta oikolukuehdotuksesta. Kuvassa ylhäällä hakukentässä on käyttäjän väärin kirjoittama hakutermi "oikluku" ja alhaalla vasemmalla on hakukoneen ehdottama oikoluettu hakutermi "oikoluku".

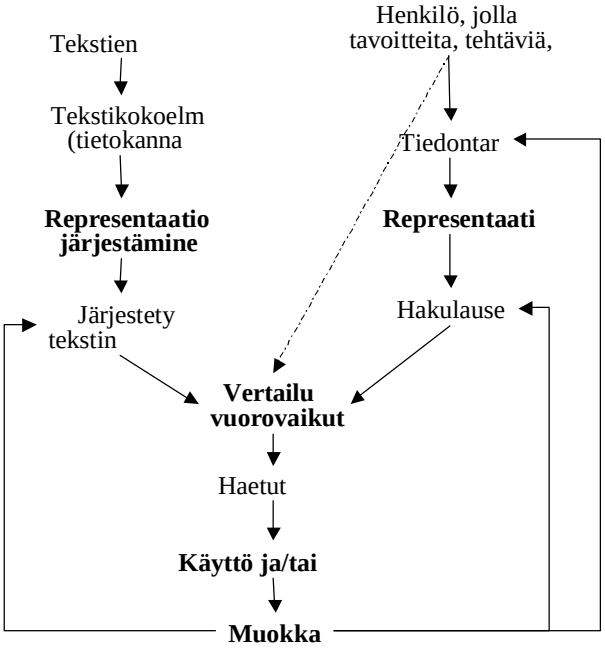
3.3.2. Passiivinen haku

Passiivisella haulla tarkoitetaan menetelmiä, joissa hakukone jonkin käyttäjän mallin perusteella automaattisesti tarjoaa käyttäjälle kohdennettua informaatiota. Passiivisia hakumenetelmiä kutsutaan myös personointiteknologioiksi, koska niissä samasta tietomassasta profiilien perusteella tuotetaan käyttäjille personoitua informaatiosisältöä. Yleisimpiä passiivisen haun menetelmiä on automaattinen niin sanottujen agenttiohjelmien tekemä tiedon suodattaminen, sekä käyttäjäprofiileihin perustuva tiedon reitittäminen.

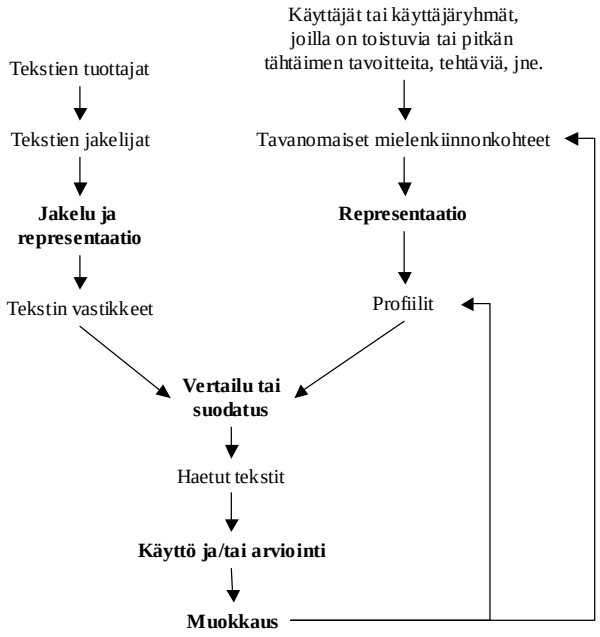
Tiedon suodatuksella tarkoitetaan informaation valintaa ennalta määriteltyjen kriteerien perusteella. Ihminen tekee jatkuvasti informaation valikointia mm. valitsemalla minkä sanomalehden ostaa ja mitä artikkeleita sieltä lukee. Saatavilla olevan informaation määrän kasvaessa on nähty tarve kehittää tietokonepohjaista automaattista valikointia informaatiotulvan kontrolloimiseksi. Eräs esimerkki agenttipohjaisesta automaattisesta tiedon suodatuksesta on roskapostin suodatus. (Foltz, Dumais 1992)

Reititys perustuu eksplisiittisiin käyttäjäkohtaisiin profiileihin. Profiili on joukko hakusääntöjä, jotka käyttäjä on tallentanut etukäteen. Hakukoneen kannalta profiilit toimivat kuten tallennetut hakulausekkeet. Reitityksessä hakukone vertailee jokaista uutta indeksiin tallennettavaa tiedostoa kaikkiin ennalta tallennettuihin

hakulausekkeisiin eli käyttäjien profiileihin ja lähettää reaaliaikaisesti tai lähes reaaliaikaisesti ilmoituksen sovellukselle, mikäli uusi tiedosto vastaa käyttäjän hakupreferenssejä. Systeemin toimintaa voidaan optimoida vertaamalla uusia indeksiin tallennettuja tiedostoja tallennettuihin profiileihin ennalta määrätyn aikataulun mukaisesti. Tällöin reaaliaikaisuus ei täysin toteudu, mutta järjestelmän kapasiteettivaatimukset ovat kevyemmät. Uusien tiedostojen automatisoitu vertaaminen talletettuihin profiileihin mahdollistaa reaaliaikaisen käyttäjän preferenssien mukaan suodatetun sisältövirran tarjoamisen. Kuvissa 7 ja 8 on esitetty yleinen malli aktiiviselle haulle ja passiiviselle tiedon reitittämiselle.



Kuva 7. Yleinen malli tiedon aktiiviselle haulle. Muokattu (Belkin, Croft 1992)



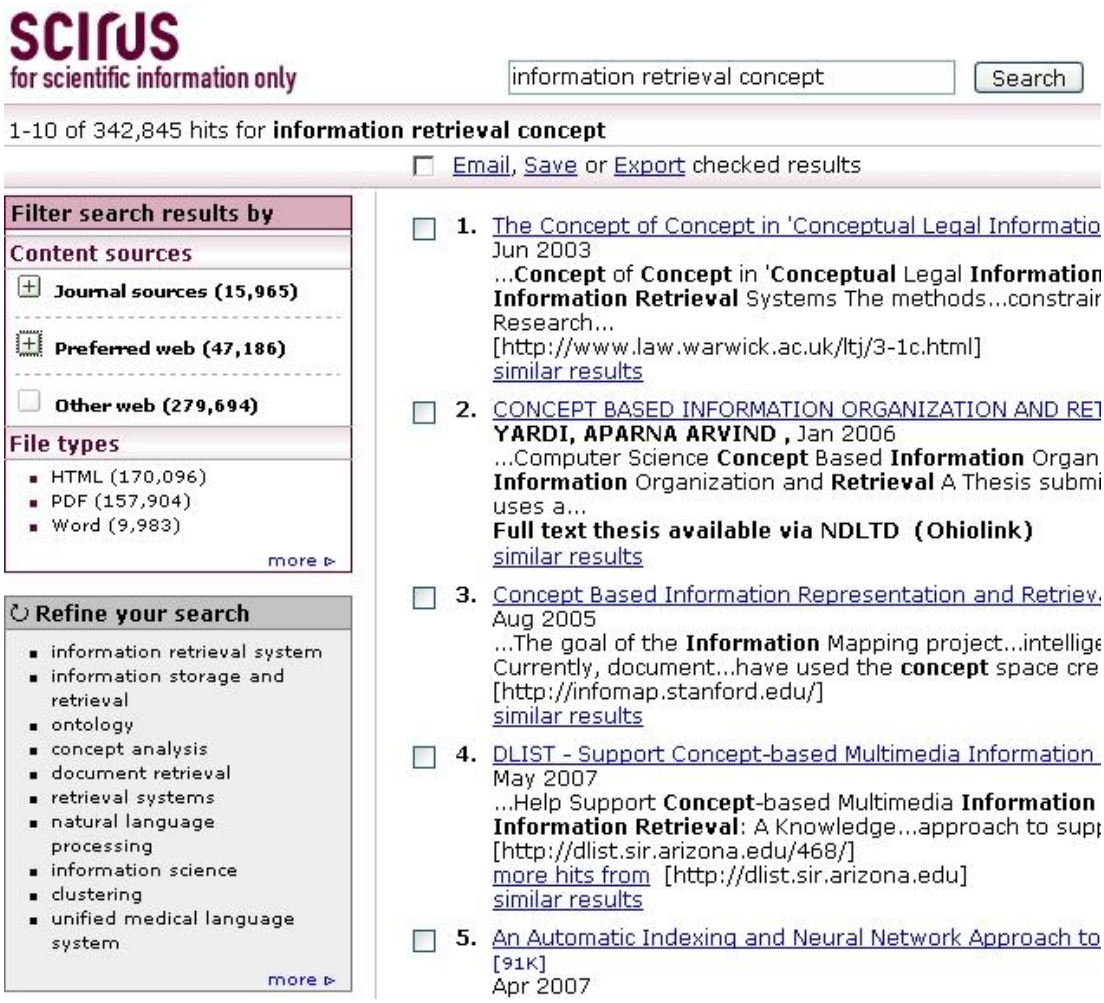
Kuva 8. Yleinen malli tiedon suodattamiselle. Muokattu (Belkin, Croft 1992)

Jos esimerkiksi yritys tai muu organisaatio ostaa uutispalveluna kaikki mahdolliset uutiset, jotka se sitten tarjoaa luettavaksi työntekijöilleen, syntyy helposti tilanne, että kaikki työntekijät saavat liikaa uutisia, jotka eivät kiinnosta heitä. Reitittäminen tarkoittaa, että työntekijät tallentavat uutispalvelun käyttäjäprofiileihinsa pitkäaikaisia hakulausekkeita, jotka kuvaavat heidän kiinnostuksenkohteitaan. Palvelin testaa näitä profiileihin tallennettuja hakulausekkeita jatkuvasti saapuvaa dokumenttivirtaa kohtaan ja välittää (reitittää) vain hakuehdon täyttävät dokumentit kullekin työntekijälle. (Schütze, Hull & Pedersen 1995) (Belew 2000)

3.3.3. Vuorovaikutteinen haku

Vuorovaikutteisia hakumenetelmiä ovat mm. käsitehaku (concept search), samankaltaisten haku ja automaattinen relevanssipalaute (relevance feedback).

Käsitehaku hyödyntää laskennallisia tekniikoita sanojen merkitysten esittämiseen dokumenteissa. Sanojen merkityksiä tulkitaan sanojenvälisiä yhteyksiä tutkimalla. Käsitehaussa etsintä kohdistuu yksittäisten sanojen sijasta säännönmukaisuuksiin eri sanojen esiintymisyhteyksissä. Esimerkiksi sana "malli" saattaa esiintyä usein muodista kertovissa uutisissa sanojen "vaate" ja "muoti" yhteydessä, mutta myös tiedeuutisissa sanan "matemaattinen" yhteydessä. Eksaktista asiasanahausta poiketen käsitehaku pyrkii huomioimaan, miten eri ihmiset ilmaisevat saman idean eri sanoilla (Deniston 2003). Käsitehaussa tulokset ryhmitellään ja esitetään käyttäjälle näiden apusanojen kanssa, jolloin hän voi valita tarkoittamansa käsitteen ja rajata alkuperäistä hakua. Tulosten ryhmittelyssä voidaan käyttää vastaavanlaista klusterointitekniikkaa, kuin sisällönkuvailussakin.



Kuva 9. Esimerkki tieteellisen tiedon hakuun tarkoitetulta scirus.com sivustolta löytyvästä hakupalvelusta, joka hyödyntää käsitehakua. Kuvassa ylhäällä hakukentässä on alkuperäiset käyttäjän antamat hakutermit: information, retrieval ja concept. Alhaalla vasemmalla on ”Refine your search” laatikko, jossa hakukone ehdottaa aihealueita, joissa käyttäjän antamilla termeillä on hieman toisistaan eroavia merkityksiä.

Samankaltaisten haulla tarkoitettaan vuorovaikutteista menetelmää, jossa käyttäjä itse toteaa jonkun dokumentin relevantiksi ja tätä dokumenttia kokonaisuutenaan käytetään hakulausekkeena ja etsitään lisää samankaltaisia.

Automaattinen relevanssipalaute on vuorovaikutteinen menetelmä hakulausekkeiden uudelleenmuotoiluun tiedonhakijan antaman palautteen avulla. Hakija arvioi, mitkä dokumenteista ovat relevantteja ja mitkä epärelevantteja. Tämän relevanssipalautteen avulla muotoillaan automaattisesti uusi kysely. Ideana on rakentaa sellaisia uusia kyselyjä, jotka tuottavat enemmän relevanttien kaltaisia dokumentteja ja vähemmän epärelevanttien kaltaisia dokumentteja. (Järvelin 1995)

3.3.4. Hakutulosten esittäminen

Hakutuloksia esitettäessä pitäisi käyttäjälle esittää kaikkein relevanteimmat tulokset ensin ja antaa mahdollisuudet itse tiivistelmien tai poimintojen avulla arvioida dokumenttien relevanttiutta ja edelleen tarjota työkaluja haun parantamiseen.

Hakutuloksia esitettäessä näytetään yleensä listaus, jossa on dokumentin otsikko ja lyhyt muutaman rivin tiivistelmä dokumentin sisällöstä. Tiivistelmiä tai poiminta

voidaan tuottaa automaattisesti hakulausekkeiden perusteella, jolloin niistä näkyy ne osat dokumentin sisältämästä tekstistä, jotka parhaiten vastaavat hakulauseketta.

Kuva 10. Google hakukoneen ensimmäiset hakutulokset hakulausekkeelle "uutispalvelut AND hakuteknologiat", kunkin hakutuloksen kohdalla on esitetty poiminta, jossa hakusanat on lihavoitu ja niitä ympäröivää tekstiä on näytetty vähän.

4. Uutispalvelut

Informaatioyhteiskunnassa tiedon saannin voidaan sanoa olevan perustarve. Uutisten välityksessä eri toimijoilla on omia palveluitaan. Kuluttajat saavat uutiset yleensä lehdistön, television, radion ja Internetissä toimivien medioiden kautta. Näitä medioita voidaan kuvata uutisoinnin vähittäismyyjiksi, koska he ovat suoraan yhteydessä loppuasiakkaisiin. Perinteisesti uutisten tukkukauppiaita ovat olleet uutistoimistot. Niiden tarkoituksena on kerätä uutisia ja välittää niitä muille vähittäismyyntimedioille. Uutistoimistoja käsitellään kappaleessa 4.1 ja uutistoimistojen tarjoamia palveluita kappaleessa 4.2. Tämän tutkimuksen kohdeyritys Suomen Tietotoimisto esitellään kappaleessa 4.3.

4.1. Uutistoimistot

Ranskalaisen AFP uutistoimiston edeltäjää, 1832 perustettua Havasta, pidetään maailman ensimmäisenä uutistoimistona. Suurimmassa osassa Euroopan valtioita toimi oma kansallinen uutistoimisto jo 1800-luvun loppuun mennessä. Myös Suomeen perustettiin uutistoimisto, STT, jo ennen itsenäistymistä vuonna 1887. Uutistoimistojen voidaan sanoa olleen ensimmäisiä näkyviä ilmentymiä globalisaatiosta ja alusta saakka niiden roolina on ollut linkittää eri valtiota ja valtiontalouksia toisiinsa välittämällä tietoa yli maantieteellisten- ja kielirajojen. Ennen ensimmäisten uutistoimistojen perustamista lehtien ulkomaan uutiset perustuivat pääosin ulkomaalaisten lehtijuttujen kääntämiseen ja lainaamiseen, sekä harvoissa tapauksissa omien ulkomaankirjeenvaihtajien käyttöön. Kansainvälisyyden ohella uutistoimistoilla on merkittävä rooli myös kansallisen identiteetin luomisessa ja usein ne ovatkin olleet ensimmäisten organisaatioiden joukossa, joita itsenäistyneisiin valtioihin on perustettu. (Boyd-Barrett, Rantanen 2002)

Nykyään toimivat uutistoimistot voidaan jakaa kansallisiin uutistoimistoihin, joita on suuressa osassa itsenäisiä valtioita, sekä muutamaan selvästi monikansalliseen uutistoimistoon. Merkittävimmät kansainväliset uutistoimistot ovat AFP (Ranska), AP (USA) ja Reuters (UK), joilla kaikilla on myös perinteikäs ja merkityksellinen kansallinen identiteetti. Kolmesta suuresta uutistoimistosta ainoastaan Reuters on pörssinoteerattu yhtiö kahden muun ollessa, useiden kansallisten uutistoimistojen tapaan, asiakkaidensa eli yksittäisten sanomalehtien ja mediatalojen omistamia. Kansainvälisesti merkittävien uutistoimistojen määrä on vähentynyt samanaikaisesti kun suuret uutisten vähittäismyyntiin keskittyneet mediatalot, kuten BBC ja CNN ovat perustaneet omia riippumattomia verkostojansa kansainväliseen uutisten hankintaan. (Boyd-Barrett, Rantanen 2002)

Kansalliset uutistoimistot ovat useinmiten johtavia uutisten kerääjiä yhden maan sisällä. Ne yhdistävät maanlaajuiset, alueelliset ja paikalliset mediat verkostoksi, jonka keskellä kansallinen uutistoimisto toimii uutisten tukkujakelijana. Tunnettuja kansallisia uutistoimistoja, joilla on myös omaa ulkomaanuutisten hankintaa on mm. Press Association Britanniassa, dpa Saksassa, EFE Espanjassa ja ANSA Italiassa. Kansalliset uutistoimistot toimivat myös linkkinä muiden maiden uutistoimistojen ja kansainvälisten uutistoimistojen suuntaan, jolloin kiinnostavat uutiset leviävät maailmalle kansallisen uutistoimiston kautta. Kansainväliset uutistoimistot puolestaan välittävät ulkomaanuutisia kansallisille uutistoimistoille ja siten paikalliset mediat tulevat ottaneeksi suuren osan ulkomaanuutisista joko suoraan tai

epäsuorasti muutaman kansainvälisen uutistoimiston tarjonnasta. Osa kansallisista uutistoimistoista on perustettu kansainvälisten toimistojen avustamana. (Boyd-Barrett, Rantanen 2002)

Uutisvälityksen luonteen mukaisesti niin kansalliset, kuin kansainvälisetkin uutistoimistot ovat vahvasti verkottuneita toisiinsa ja ne välittävät joko suoraan tai muokattuina toistensa tuottamia uutisia alueilta, joita heidän oma uutistenhankintansa ei kata. Euroopassa uutistoimistojen kattojärjestönä toimii European Alliance of News Agencies EANA. Monet uutistoimistot ovat myös osakkaana uutisvälityksen standardeja suunnittelevassa ja ylläpitävässä International Press Telecommunications Councilissa IPTC:ssä.

4.2. Uutistoimistojen tarjoamat palvelut

Liiketoimintana uutisten välitys on uutisjuttujen tekemistä ajankohtaisista tapahtumista ja niiden välittämistä asiakkaille. Kuluttajat eivät halua saada kaikkia heidän näkökulmastaan hyödyttömiä uutisia, vaan tietoa heitä kiinnostavista tapahtumista. Mediakenttä on jakautunut lukuisiin toimijoihin, jotka pyrkivät palvelemaan omaa kohderyhmäänsä mahdollisimman hyvin tarjoamalla relevantteja ja ajankohtaisia uutisia. Perinteisesti uutistoimistot eivät ole tarjonneet uutisia suoraan kuluttajille, vaan ovat keskittyneet palvelemaan lehdistöä ja muita medioita ja uutistoimistojen palveluita on yleensä käyttäneet toimittajat.

Klassinen jako uutisten tukkukauppiaisiin ja vähittäismyyjiin on edelleen suurelta osin totuudenmukainen, mutta rajat eri toimintaroolien välillä ovat hälvenemässä. Uutistoimistot ovat entistä enemmän hankkineet myös median ulkopuolisia asiakkaita. Median ulkopuolelta ensimmäisenä uutistoimistojen palveluita ovat ostaneet finanssialan toimijat, kuten pörssimeklarit, joille tiedonvälityksen nopeus on ensisijaisen tärkeää. Myös tavalliset kansalaiset pääsevät nykyään eri medioiden tarjoamien internetpalveluiden kautta lukemaan nopeasti uutistoimistojen välittämiä uutisia, joita ei ole editoitu. Näin ollen entistä suuremmalla todennäköisyydellä uutistoimiston välittämän uutisen lukee joku muu, kuin toimittaja. Vastaavasti myös yksittäisten sanomalehtien, TV-kanavien ja mediatalojen oma uutistoimistoista riippumaton uutisten hankinta on yleistynyt myös ennen vahvasti uutistoimistojen käsissä olleiden ulkomaanuutisten kohdalla. (Boyd-Barrett, Rantanen 2002)

4.3. Suomen Tietotoimisto

Vuonna 1887 perustettu Suomen Tietotoimisto on perinteikäs ja luotettu uutislähde Suomessa. STT:n toiminnan ydin on jatkuvasti päivystävä uutistoimitus, joka toimittaa omaa materiaalia päivittäin noin tuhat dokumenttia sisältäen suomen- ja ruotsinkielisen uutistarjonnan. Perusuutispalvelun oheen on kehittynyt erilaisia täydentäviä media- ja viestintäpalveluita.

Uutistoimistolla on kahdeksan aluetoimitusta Suomessa, oma edustus Brysselissä, Moskovassa, Tallinnassa, Tukholmassa ja Washingtonissa sekä useita vakituisia avustajia Suomessa ja ulkomailla. STT:n palveluksessa on noin 130 ihmistä, joista 110 toimittajaa. Kansainvälisessä uutisvälityksessä STT toimii tiiviisti osana uutistoimistojen verkostoa. STT:n tärkeimpiä kumppaneita ovat johtavat kansainväliset uutistoimistot. STT käyttää eniten brittiläisen Reutersin, ranskalaisen AFP:n, saksalaisen DPA:n ja yhdysvaltalaisen AP:n materiaalia. Vastavuoroisesti lukuisilla uutistoimistoilla on käytössään STT:n palvelut.(STT 2006)

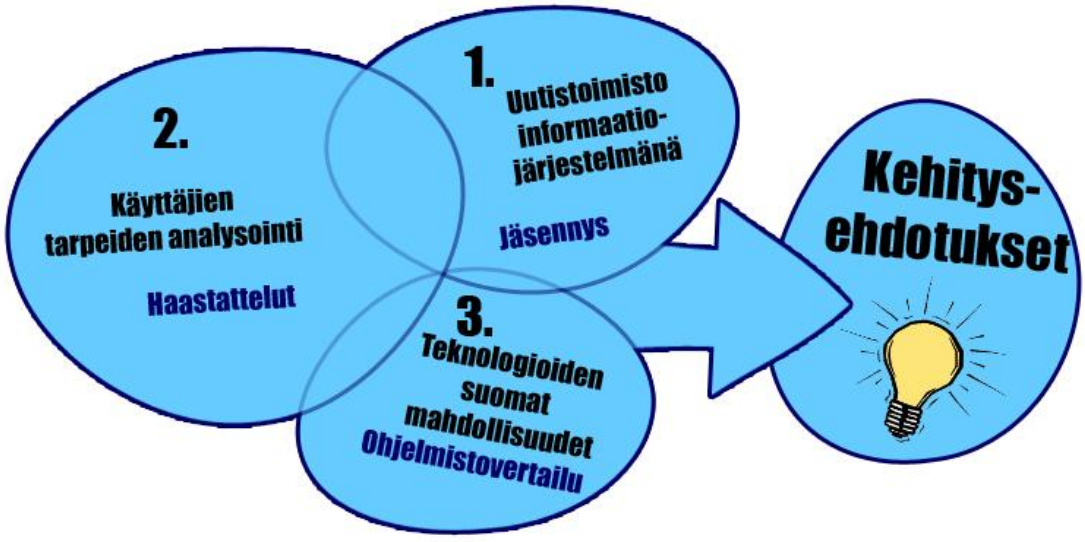
STT on media-alan yritysten omistama osakeyhtiö, kuten monien muidenkin maiden kansalliset uutistoimistot. Suurimmat omistajat ovat Alma Media (28,2 %), Sanoma Osakeyhtiö (22,1 %) ja TS-Yhtymä (21,0 %). Eri mediat aina pienistä maakuntalehdistä valtakunnallisiin sähköisiin medioihin ovat myös STT:n suurin asiakasryhmä. Media-asiakkaiden lisäksi STT toimittaa uutisia myös muille tahoille mm. järjestöille, yrityksille, valtionhallinnolle ja tutkimuslaitoksille. (STT 2006)

5. Tutkimusasetelma

Tutkimuksen lähtökohtana on, että STT:n journalistisesti tuottamasta uutismateriaalista on mahdollista jalostaa uusia entistä paremmin käyttäjien tarpeita vastaavia uutispalveluita hyödyntämällä haku- ja personointiteknologioita. Tavoitteena on antaa perusteltuja kehitysehdotuksia, joilla STT:n palvelutarjontaa voidaan parantaa.

Tulevaisuudessa STT:n asiakaskunnan kasvun odotetaan tulevan todennäköisimmin median ulkopuolelta mm. yrityksistä, järjestöistä ja julkisen hallinnon alueelta. Kehitteillä olevat uudet palvelut on suunnattu ensisijaisesti median ulkopuolisille organisaatioille. Näissä organisaatioissa uutispalveluiden loppukäyttäjät ovat yleensä tiedotuksesta vastaavaa henkilökuntaa, hallintohenkilökuntaa tai tutkijoita. Tutkimuksen ja myös kehitettävien palveluiden kohderyhmä on median ulkopuolisissa organisaatioissa työskentelevät työssään uutispalveluita käyttävät henkilöt.

Luonteva aloitus tutkimukselle oli STT:n palveluiden nykytilan analysointi. Nykytilannetta analysoitiin jäsentämällä uutistoimiston toimintaa informaatiojärjestelmänä ja kartoittamalla käyttäjien tarpeita haastatteluilla. Hakuteknologioiden suomia mahdollisuuksia selvitettiin ohjelmistovertailun avulla. Lopuksi esitetään kehitysehdotukset STT:n informaatiojärjestelmään, jotka pohjautuvat käyttäjien tarpeisiin ja teknologian suomiin mahdollisuuksiin. Tutkimuksen eteneminen nykytilan analysoinnista kehitysehdotuksiin näkyy kuvassa 11.



Kuva 11. Tutkimuksen vaiheet; nykytilaa hahmotetaan käyttäjien tarpeiden, teknologian mahdollisuuksien ja informaatiojärjestelmän mallintamisen kautta ja niiden pohjalta tehdään perusteltuja parannusehdotuksia.

Uutistoimistot ovat vähän tutkittu, mutta informaatiotutkimuksen kannalta erittäin mielenkiintoinen kohde. Tutkimuskysymykset määritellään kappaleessa 5.1 ja tutkimusaineistot ja käytetyt menetelmät esitellään kappaleessa 5.2.

5.1. Tutkimuskysymykset ja tutkimuksen vaiheet

Uutistoimiston jatkuvasti päivittyvissä tietokannoissa ja arkistossa on kattavasti monia eri toimijoita kiinnostavaa informaatiota. Tästä aineistosta halutaan tuottaa eriytettyjä uutispalveluita modernin teknologian avulla mahdollisimman joustavasti erilaisille toimijoille. Reaaliaikaiset luokittelu- ja hakuteknologiat sekä personointiteknologiat ovat kehittyneet viimevuosina niin, että niiden varaan on mahdollista rakentaa kustannustehokkaasti toimivia uusia palveluita. Jatkosuunnittelun tueksi halutaan selvittää, minkälaisia palveluita voidaan tuottaa ja mille palveluille on erityisesti kysyntää.

Tutkimusaiheen ongelmakentästä erityisiksi tutkimuskysymyksiksi nostettiin seuraavat:

1.) Miten uutistoimistojen ja muiden uutislähteiden nykyisiä palveluita käytetään kohderyhmän arkityössä ja minkälaisia parannustarpeita niihin liittyy?

2.) Minkälaisia haku- ja personointiteknologioita on tarjolla ja minkälaisia ominaisuuksia kaupallisilla ohjelmistoilla on?

3.) Miten haku- ja personointiteknologioilla voidaan tuottaa uutispalveluja helpottamaan loppukäyttäjien työtä asiakasorganisaatioissa?

5.1.1. Tutkimuksen vaiheet

Käytännössä tutkimus jakautui neljään osa-alueeseen, jotka tehtiin toisiaan tukien osittain samanaikaisesti. Kaikki osa-alueet eivät olleet tutkimuksessa kuitenkaan esillä yhtä suurilla painoarvoilla, vaan käyttäjien tarpeiden analysointiin ja kehitysehdotusten tekemiseen käytettiin huomattavasti enemmän aikaa, kun taas ohjelmistovertailu ja uutistoimiston informaatiojärjestelmän jäsennys tehtiin kevyemmin.

STT:n informaatiojärjestelmän mallintaminen. Ensimmäisessä vaiheessa mallinnettiin STT:n informaatiojärjestelmän nykyistä toimintaa Ingwersenin ja Järvelinin kirjassa: *"The Turn, Integration of Information Seeking and Retrieval in Context"* (Ingwersen, Järvelin 2005) esittelemää viitekehystä hyväksikäyttäen.

1. **Käyttäjänäkökulman tutkiminen teemahaastatteluiden avulla.** Toisessa vaiheessa tutkittiin niitä prosesseja, joissa STT:n nykyisiä palveluita ja muita uutislähteitä käytetään median ulkopuolisissa STT:n asiakasorganisaatioissa. Teemahaastattelun muodossa selvitettiin kahdeksalta työssään STT:n nykyisiä palveluita ja muita uutislähteitä käyttävältä henkilöltä, miten ja mihin he käyttävät uutislähteitä ja minkälaisia parannustarpeita uutisten seurannan työkaluihin kohdistuu. Haastattelut olivat kestoltaan 40–60 minuuttia, ja ne tehtiin aina asiakasorganisaation tiloissa. Haastattelumateriaali täydensi ensimmäisessä vaiheessa luotua STT:n informaatiojärjestelmän mallia käyttäjien näkökulmasta.
2. **Teknologian suomien mahdollisuuksien kartoitus ohjelmistovertailun avulla.** Kolmannessa vaiheessa tutustuttiin markkinoilla oleviin haku- ja personointiteknologioihin vertailemalla uutisaineiston luokitteluun soveltuvia ohjelmistoja keskenään. Tämän vaiheen yhteydessä selvitettiin STT:n sisällä niitä tarpeita, joihin haku- ja personointiteknologioiden odotettiin vastaavan. Viiden kaupallisen ohjelmistojen ominaisuuksia verrattiin löydettyihin tarpeisiin. Lopuksi testattiin neljän potentiaalisimman ohjelmistontoimittajan

uutisten luokitteluominaisuuden toimivuutta STT:n omalla uutisaineistolla. Ohjelmistovertailu antoi teknisestä näkökulmasta tietoa STT:n informaatiojärjestelmän kehittämiseen.

3. **Kehitysehdotusten tekeminen.** Viimeisessä vaiheessa tehtiin kehitysehdotuksia, joissa markkinoilla olevia haku- ja personointiteknologioita hyödyntämällä vastataan käyttäjien haastatteluista esiin nousseisiin tarpeisiin. Otin mallia kehitysehdotuksiin mm. muiden kansallisten uutistoimistojen toteuttamista palveluista, hakuteknologioita kehittävien yritysten toteuttamista sovelluksista ja osin ideoin täysin puhtaalta pöydältä.

5.2. Tutkimusaineistot ja menetelmät

Tässä kappaleessa esitellään tutkimusmenetelmät ja käytetyt aineistot kahdesta toisiaan täydentävästä tutkimuksesta. Työssä on sovellettu sekä määrällisiä, että laadullisia menetelmiä. Uutispalveluiden loppukäyttäjien tarpeita on selvitetty laadullisin menetelmin haastattelututkimuksena ja ohjelmistovertailussa on käytetty sekä laadullisia että määrällisiä menetelmiä.

Kvantitatiivisella eli määrällisellä tutkimuksella tarkoitetaan yleensä tilastolliseen analyysiin perustuvaa tutkimusta. Kvalitatiivisella eli laadullisella tutkimuksella tarkoitetaan tulkinnallisen tutkimuksen tekemistä. Kvalitatiivisessa tutkimuksessa yleisesti käytettyjä metodeita ovat mm. haastattelu, havainnointi ja tekstianalyysi.(Metsämuuronen 2003)

Määrällinen tutkimus mahdollistaa tapahtumien yleisluontoisen jakautumisen selvittämisen ja tilastollisen yleistettävyyden, mutta se edellyttää ilmiöiden rajaamista ja järjestämistä kokeiksi, joissa vaikuttavat tekijät on tarkkaan kontrolloituja. Laadullisella tutkimuksella voidaan paremmin selvittää tapahtumien yksityiskohtaisia rakenteita ja toisaalta se soveltuu luonnollisten tilanteiden tutkimiseen, joissa kaikkia vaikuttavia tekijöitä ei pyritäkään kontrolloimaan. (Syrjälä et al. 1994)

Käyttäjäkeskeistä näkökulmaa valotetaan käyttäjien tarpeiden analysoinnilla, jonka tutkimusmenetelmät esitellään kappaleessa 5.2.1. Järjestelmäkeskeisempää ajattelua edustaa markkinoilla olevien kaupallisten uutisaineiston luokitteluun soveltuvien ohjelmistojen vertailu, jossa käytetyt aineistot ja menetelmät esitellään kappaleessa 5.2.3.

5.2.1. Käyttäjähaastattelut

Tässä työssä haastattelumenetelmänä käytettiin teemahaastattelua, jossa oli ennakkoon määritellyt aihealueet ja niitä tukevat apukysymykset. Haastattelumenetelmänä teemahaastattelu on kontrolloidun lomakehaastattelun ja avoimen haastattelun välimuoto. Lomakehaastattelussa haastattelu tapahtuu ennalta suunnitellun lomakkeen mukaisesti kun taas avoimessa haastattelussa ei ole lainkaan ennalta valmisteltua runkoa, vaan haastattelu etenee keskusteluna haastattelijan ohjatessa keskustelua. (Hirsjärvi, Hurme 1988)

Haastattelut käytiin valittujen teemojen pohjalta vapaamuotoisena keskusteluna, jotka nauhoitettiin. Aikaa haastatteluun varattiin etukäteen noin yksi tunti haastateltavaa kohden ja haastattelut tehtiin haastateltavan organisaation tiloissa.

Seuraavaksi esitellään haastattelun kohderyhmä, haastatteluteemat, ja haastatteluaineiston analyysimenetelmät.

Haastattelututkimuksen kohderyhmä

Haastattelututkimuksen tarkoituksena on selvittää, minkälaisia tarpeita on median ulkopuolisilla uutispalveluiden käyttäjillä. Haastattelututkimus rajataan koskemaan ainoastaan STT:n nykyisiä median ulkopuolisia asiakkaita, sillä kaikkiaan median ulkopuolisia organisaatioita, jotka ovat STT:n potentiaalista asiakaskuntaa on paljon ja ne ovat tutkimuksen kannalta hankalasti tavoitettavia.

Haastatteluilla selvitettiin, mihin tarkoituksiin ja millä tavalla median ulkopuoliset asiakkaat käyttävät STT:n nykyisiä palveluita. Tutkittavien perusjoukon muodostavat asiakasorganisaatioiden sisällä toimivat palveluiden loppukäyttäjät, jotka ovat yleensä tiedotuksesta vastaavaa henkilökuntaa, hallintohenkilökuntaa ja tutkijoita.

Kohderyhmästä valittiin haastateltaviksi asiakkaita, jotka käyttävät mahdollisimman monipuolisesti STT:n nykyisiä palveluita ja jotka edustavat hieman erilaisia organisaatioita. Kahdeksan haastatellun joukossa oli niin arkiston, tiedotepalveluiden, tekstiviestiuutisten, toimialaseurannan, grafiikan kuin listojen ja tapahtumakalenterinkin käyttäjiä (katso STT:n palvelut taulukko 6). Lisäksi kaikki haastatellut käyttivät säännöllisesti perus uutispalvelua. Haastatellut olivat henkilöitä, jotka käyttivät itse uutispalveluita työssään. Joissain tapauksissa haastatellut olivat myös vastuussa STT:n ja muiden uutispalveluiden ostamisesta oman organisaationsa ja itsensä käyttöön. Tutkimukseen valitut organisaatiot ovat tutkimuksen tilaajan tiedossa, mutta niitä ei yksilöidä tutkimuksen tulosten yhteydessä ja etenkin kaikki tutkimukseen osallistuneet loppukäyttäjät pidetään ehdottomasti anonyymeinä.

Haastateltavista kuusi oli miehiä ja kaksi naisia. He edustivat yksityisiä yrityksiä (3 kpl.), järjestöjä (3 kpl.) ja julkishallinnon organisaatioita (2 kpl.) Haastatellut olivat joko johtavassa asemassa itse (2 kpl.), vastuussa organisaationsa viestinnästä (4 kpl.) tai tekivät työkseen tiedonhakua (2 kpl.).

Haastatteluteemat

Haastattelututkimuksen tavoitteena oli selvittää, miten uutistoimistojen ja muiden uutislähteiden nykyisiä palveluita käytetään kohderyhmän arkityössä ja minkälaisia parannustarpeita niihin liittyy? Tästä tutkimuskysymyksestä muodostettiin teemahaastatteluun neljä teemaa: palvelut, käyttötarkoitukset, käyttötilanteet ja kehityskohteet. Haastatteluteemat ja niihin liittyvät apukysymykset on esitelty taulukossa 3.

Taulukko 3. Teemahaastatteluissa käytetyt teemat.

Teema 1: Palvelut **(Mitä?)**	Keskeinen teema haastatteluissa on informaatiopalvelut yleensä ja erityisesti STT:n nykyisellään tarjoamat palvelut. Kaikki muut haastatteluteemat sivuavat ja syventävät palveluteemaa. Pyritään selvittämään, mitä STT:n nykyisiä palveluita ja muita vastaavia kilpailevia tai täydentäviä informaatiopalveluita haastateltavat käyttävät työssään. Palveluiden ohella selvitetään yleisemmin, minkälaisia tietolähteitä (STT ja muut, sähköiset, henkilö, yms.) käytetään, miten niiden luotettavuutta arvioidaan ja miten lähteitä valitaan.
Teema 2: Käyttötarkoitukset (Miksi?)	Mihin tarkoituksiin organisaatiossa käytetään STT:n palveluita ja muita vastaavia palveluita? Miksi joku tietty palvelu on käytössä? Minkälaista tietoa STT:n ja muiden palveluiden kautta yleensä haetaan? Mihin muuhun toimintaan haastateltavan työssä STT:n palveluiden käyttö liittyy? Pyritään selvittämään perimmäisiä syitä siihen, miksi palveluita käytetään. STT:n uutispalvelua epäilemättä käytetään uutisten seuraamiseen, mutta miksi haastateltava seuraa uutisia? Esimerkkejä työtehtävistä, joissa STT:n palveluja ja muita käytetään.
Teema 3: Käyttötilanteet (Miten?)	Taustatiedoksi selvitetään, mikä on haastateltavan rooli organisaatiossa ja pyritään hahmottamaan yleiskuva organisaation tehtävistä. Pyydetään haastateltavaa kuvailemaan, mitä hän konkreettisesti tekee työkseen ja mikä on hänen ammattitaustansa. Selvitetään haastateltavan STT:n palveluihin liittyviä käyttörutiineja, toimintatapoja, aikaa, paikkaa, kestoa, säännöllisyyttä, frekvenssiä, ympäristöä, kontekstia, yhteistyötä ja muiden henkilöiden osallistumista. Mikä on tiedonhankinnan prosessi, lähtötilanne, päämäärät, mitä tehdään ja toisaalta mikä on työn lopputulos? Pyydetään myös konkreettisia esimerkkejä hakutehtävistä.
Teema 4: Kehityskohteet	Mikä on hyvää ja mikä huonoa STT:n nykyisissä palveluissa? Kysytään haastateltavilta konkreettisia esimerkkejä ongelmatilanteista ja selvitetään palveluihin liittyviä parannustarpeita. Kysytään palvelukohtaisia kommentteja ja vertailuja muihin vastaaviin palveluihin. Selvitetään yleisellä tasolla tiedonhakuun ja kulkuun liittyviä parannustarpeita.

Haastatteluaineiston luokittelu ja analysointi

Nauhoitetut haastattelut litteroitiin haastateltavien osalta sanatarkasti. Kestoltaan toteutuneet haastattelut olivat puolesta tunnista tuntiin.

Litteroitua haastatteluaineistoa luettiin läpi ja kaikki siinä esiintyvät sisällöltään merkitykselliset kommentit luokiteltiin Atlas[1] ohjelmistolla. Kahdeksasta haastattelusta poimittiin yhteensä 356 kommenttia, jotka luokiteltiin 29 luokkaan. Aineistoa luokiteltaessa sama kommentti luokiteltiin aina yhteen tai useampaan luokkaan, esimerkiksi arkiston käytettävyyteen liittyvä kommentti luokiteltiin "Käytettävyys" ja "Arkisto" luokkiin.

Luokittelun avulla pyrittiin korostamaan haastatteluissa todellisuudessa esille nousseita aiheita ja muokkaamaan aineistoa tulkinnan kannalta helpommin käsiteltävään muotoon. Luokkajakoa ei määrätty ennakolta, vaan luokkajako muokkautui aineiston käsittelyn edetessä, joitain luokkia yhdistettiin ja tarpeen mukaan uusia luokkia luotiin. Loppujen lopuksi käytetyt luokat noudattavat löyhästi haastattelun ennalta valittuja teemoja. Käytetyt luokat ja niihin kertyneiden kommenttien lukumäärät on esitetty liitteessä A.

Koska sama henkilö teki haastattelut, litteroinnin ja luokittelun suhteellisen lyhyessä ajassa oli mahdollista aloittaa aineiston analysointi jo ennen luokittelun valmistumista. Valittu luokkajaon ohjasi myöhempää analyysia, joka tapahtui lukemalla luokkien mukaan järjestettyjä kommentteja läpi ja etsimällä niistä yleislinjoja ja usein toistuvia merkittäviä samaa asiaa käsitteleviä kommentteja.

Haastattelututkimus oli puhtaasti laadullinen tutkimus, joten en ole tehnyt mitään johtopäätöksiä pelkän luokkajaon tai esimerkiksi luokkien sisältämien kommenttien määrien perusteella. Luokituksen tarkoituksena oli ainoastaan helpottaa aineiston käsittelyä. Luokkien mukaan järjestettyinä eri haastatteluista lähtöisin olevat samaa asiaa koskevat kommentit oli luettavissa peräkkäin.

5.2.2. Ohjelmistovertailu

Ohjelmistovertailun tavoitteena oli saada selkeä kuva markkinoilla olevista haku- ja personointiteknologioista, niiden ominaisuuksista ja sovelluksista uutispalveluiden tuottamiseen. Vertailussa perehdyttiin viiden vuonna 2006 markkinoilla olleen uutisten automaattiseen luokitteluun soveltuvan kaupallisen ohjelmiston ominaisuuksiin. Järjestelmätoimittajat pidetään tässä yhteydessä anonyymeina. Järjestelmäntoimittajista kolme oli kotimaisia ja kaksi kansainvälisiä yrityksiä. Vertailluilla ohjelmistoilla oli luokittelun ohella vaihteleva määrä muitakin haku- ja personointiominaisuuksia, mutta ainoastaan luokitteluominaisuuden toimivuutta testattiin käytännössä eri ohjelmistoilla.

Tuotteiden ilmoitettuja ominaisuuksia verrattiin STT:n tarpeisiin, jonka jälkeen parhaiten tarpeisiin vastaavien ominaisuuksien kohdalla vertailtiin toteutusteknologioita keskenään ja arvioitiin niiden hyvyyttä ja kattavuutta. Lisäksi

[1] ATLAS.ti on visuaaliseen kvalitatiiviseen analysointiin suunniteltu ohjelma, jonka aineistoksi sopivat tekstit, grafiikat, äänet ja videot. http://www.atlasti.com/

yrityksistä ja tuotteista kerättiin päätöksenteon tueksi muuta taustatietoa, kuten asiakasreferenssejä ja hintatietoja. Kartoituksen pohjalta valittiin lupaavimmat luokittelujärjestelmät käytännön testeihin, joissa ohjelmistoilla luokiteltiin STT:n omaa uutisaineistoa nykyisin käytössä olevan luokittelusanaston mukaisesti. Testauksen tavoitteena oli selvittää, mikä vertailun järjestelmistä suoriutuu parhaiten tehtävästä kylmiltään ilman erityistä järjestelmän muokkausta tai hienosäätöä.

Testausvaiheessa järjestelmätoimittajille lähetettiin ensin STT:llä käytössä oleva IPTC-pohjainen luokittelusanasto, sekä opetusaineistoksi noin. 50 000 uutisartikkelia käsittävä tekstikorpus luokittelutietoineen XML-muodossa. Varsinainen testimateriaali käsitti 1000 uutisartikkelia, jotka lähetettiin järjestelmäntoimittajille ilman luokittelutietoja. Vastauksena saatuja, eri järjestelmien antamia, luokituksia verrattiin alkuperäiseen käsintehtyyn luokitukseen. Lisäksi tehtiin pienemmällä otoksella laadullinen tutkimus, jolla pyrittiin selvittämään, onko sellaisissa luokitusehdotuksissa, joita luokittelujärjestelmät ovat antaneet, mutta joita toimittajat eivät ole käsintehtyyn luokitukseen merkinneet joukossa artikkelin sisältöä vastaavia järkeviä ehdotuksia.

Seuraavaksi esitetään STT:n haku- ja personointiteknologioihin liittyvät tarpeet, STT:n nykyisin käyttämä luokittelusanasto, ohjelmistovertailun opetus- ja testiaineistot, sekä luokittelutestin vertailumenetelmät.

Luokitteluun ja tiedonhakuun liittyvät tarpeet

Ohjelmistovertailun aluksi tehtiin kartoitus niistä STT:n tarpeista, joihin personointi- ja hakuteknologioiden odotetaan vastaavan. Tarvekartoituksen pohjalta päädyttiin vertailemaan nimenomaan ohjelmistotuotteita, joilla on mahdollista toteuttaa uutisaineiston automaattista luokittelua. Vertailuun valituissa ohjelmissa oli eri määrä muitakin toimintoja, joista osa vastasi STT:n tarpeisiin ja osa oli hyödyttömiä. Lisäksi samanlaisten toimintojenkin kohdalla eri ohjelmistojen ratkaisut eroavat toisistaan pinnan alla olevan teknologian osalta. Esimerkiksi suomen kielen tuki on voitu toteuttaa puhtaasti tilastollisin menetelmin tai kattavamman suomenkielen analysoinnin avulla tai näiden menetelmien yhteensovituksella. Näin ollen on perusteltua olettaa, että eri ohjelmistojen toteutukset nimellisesti samalle toiminnalliselle ominaisuudelle, kuten automaattiselle luokittelulle eroavat toisistaan myös laadullisesti ja suorituskyvyllisesti.

Eri järjestelmäntoimittajien ratkaisut vastaavat osittain hyvinkin erilaisiin tarpeisiin, vaikka yhtenä toiminnallisuutena kaikissa on tekstiaineiston automaattinen luokittelu. STT:llä vallitsee yleisellä tasolla yhteisymmärrys siitä, mihin luokittelujärjestelmää voitaisiin käyttää, mutta erityisiä vaatimusmäärittelyjä ei ole tehty. Raportin tavoitteena on tarjota päätöksenteon tueksi kokonaisvaltainen näkemys tarjolla oleviin tuotteisiin ja niiden ominaisuuksiin, sekä vastata kysymykseen, kuinka hyvin eri tuotteet vastaavat STT:n tämänhetkisiin tarpeisiin. Kerätty aineisto on myös helposti hyödynnettävissä siinä tapauksessa, mikäli järjestelmään liittyviä tarpeita tullaan määrittämään tarkemmin.

Luokittelu sinänsä ei ole mikään itseisarvo, vaan luokittelua tehdään, jotta voidaan tuottaa laadukkaasti valikoituja uutispalveluita asiakkaiden tarpeisiin kustannustehokkaasti. Tällaisia palveluita on nykyään jonkin verran, ennen kaikkea online-palveluissa, mutta odotukset eivät ole toteutuneet läheskään sellaisina kuin kuviteltiin. Pohjimmiltaan luokittelua tehdään, jotta etsittävät jutut löytyisivät helpommin. Tämä laajentaa tarkastelua itse luokittelusta myös tiedonhaun puolelle.

Taulukko 4. Haku- ja personointiteknologioihin kohdistuvat tarpeet STT:llä.

Tarve 1: Sisään tulevan materiaalin automaattinen esikäsittely	Uutistoimittajan työajasta merkittävä osa kuluu sisään tulevan tiedon seurantaan ja relevantin materiaalin seulontaan. STT:lle tulee sisään tiedotemateriaalia: eduskunnasta, valtioneuvostolta, virastoista, viranomaisilta ja yrityksiltä. Lisäksi sisään tulee uutistoimistomateriaalia n. 5000 uutista päivässä pääosin englanniksi. Sisään tuleva materiaali on useissa formaateissa ja eri kielillä. Sama tieto tulee usein eri kanavia pitkin, tietolähteiden luotettavuudessa on eroja ja itse sisällön kiinnostavuudessa on eroja. Tarpeena on automatisoida sisään tulevan materiaalin käsittelyä siten, että tekstit luokitellaan aihepiirien mukaisesti. Näin toimittaja saa käyttöönsä juuri häntä kiinnostaviin aihepiireihin kuuluvat tekstit ja lisäksi sisällöllisesti samankaltaiset ja samaa aihetta käsittelevät tekstit tulevat valmiiksi yhteen niputettuna.
Tarve 2: Toimitetun materiaalin luokittelun automatisointi	STT:n toimittamat tekstit luokitellaan nyt käytetyn aiheluokituksen mukaisesti käsin yhteen tai useampaan noin 1300 luokasta ja lisäksi tekstien saatteisiin kirjataan alueluokitus. Käsin luokittelu on työlästä, eikä toimittaja aina tule ajatelleeksi, mihin kaikkiin luokkiin kyseinen teksti voisi kuulua. On olemassa tarve automaattiselle tekstinluokittelijalle, joka ehdottaa riittävän luotettavasti, mihin mahdollisiin luokkiin teksti kuuluu. Ehdotetuista luokista toimittaja voi sitten valita mielestään oikeat vaihtoehdot.
Tarve 3: Yritysnimien merkinnän automatisointi	Tällä hetkellä toimittajat merkitsevät yritysten nimet tekstiin käsin, mikä on työlästä ja altista inhimillisille virheille. Nimiä ei myöskään palauteta perusmuotoon, mikä vaikeuttaa tiedon hyödyntämistä. Erisnimien ja erityisesti yritysten nimien automaattinen poiminta ja merkitseminen tekstistä olisi hyödyllistä. Näiden avulla voidaan toteuttaa hienojakoisempaa uutismateriaalin valikointia ja kehittyneempiä push-palveluita, kuten tietyn henkilön tai organisaation mediaseurantaa.
Tarve 4: Muun materiaalin luokittelu	Ei ole kustannustehokasta luokitella käsin kaikkea STT:n kautta kulkevaa materiaalia, kuten muiden maiden uutistoimistojen materiaalia tai tiedotteita. Tämän luokittelemattoman materiaalin sisällyttäminen muihin uutispalveluihin taustatiedoksi olisi hyödyllistä, mutta se edellyttäisi luokittelua. Näin ollen tarvetta täysin automaattiselle luokittelulle on.

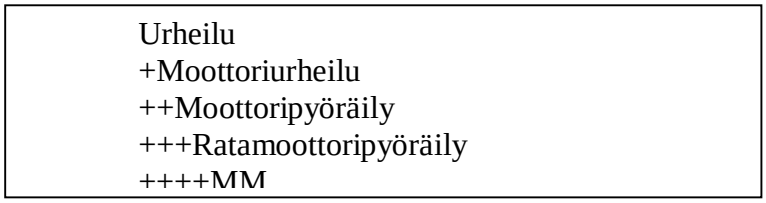
Tarve 5: Aineiston automaattinen yhdistely	STT:n asiakkaat ovat usein kiinnostuneita seuraamaan jotain tiettyä aihepiiriä useista eri näkökulmista ja haluavat siksi saada yksittäisen uutisen tai tiedotteen lisäksi myös aihetta syventävää tietoa. Toisaalta luettu uutinen saattaa herättää kiinnostuksen johonkin toiseen sivuavaan aihepiiriin. STT:llä on teoreettiset mahdollisuudet tarjota omista arkistoistaan ja muistakin tietolähteistä, kuten muiden maiden uutistoimistojen tarjonnasta sekä syventävää tietoa samasta aiheesta että materiaalia muista läheisistä aiheista. Ongelmaksi muodostuu kuitenkin tiedon yhdistely ja suhteuttaminen toisiin niin, että asiakkaalle ei tule liian laajaa tarjontaa. On tarve ratkaisulle, joka yhdistelee aineistoa ja sitä kautta mahdollistaa joko toimituksellisen työn tukemana tai automaattisesti ns. semanttiset tietopalvelut, joissa aineisto sisältää myös suhteita muuhun tietoon.
Tarve 6: Asiakaskohtaisesti eriytettyjen uutispalveluiden tuottaminen	Asiakkaiden kiinnostus eri aiheita kohtaan ja tiedon tarve ovat hyvin yksilöllisiä, joten samanlaiset tietopalvelut eivät tyydytä kaikkia asiakkaita. STT:llä on tarve tarjota asiakkaille räätälöityä täsmätietoa. Tällöin tiedon yksikön arvo kasvaa, koska sisältö on relevanttia.

Luokittelusanasto

Taksonomialla eli luokittelusanastolla tarkoitetaan sitä sanastoa, jonka mukaisiin luokkiin uutiset ryhmitellään. Kansainvälinen lehdistön ja uutistoimistojen yhteistyöjärjestö *International Press Telecommunications Council* on luonut uutisten luokitteluun standardoidun sanaston, jonka Sanomalehtien liitto on suomentanut. STT:llä on käytössään tästä suomennetusta IPTC-sanastosta STT:n tarkoituksiin muokattu sanasto, jota on mm. laajennettu huomattavasti urheilun osalta.

Käytetty sanasto käsittää kokonaisuudessaan 1293 luokkaa, jotka jakautuvat hierarkkisesti 19 pääluokkaan, ja enintään neljännelle tasolle meneviin alaluokkiin. Kolmannen ja neljännen tason alaluokat, joita on yhteensä 394 kappaletta ovat pelkästään urheilua ja kokonaisuudessaan urheiluun liittyviä luokkia on 858 kappaletta eli n. 66% sanastosta.

Sanasto toimitettiin testeissä tekstimuotoisena tiedostona, jossa on lueteltu pääluokan jälkeen sen alaluokat se. ensimmäisen tason alaluokkien edessä on + merkki ja edelleen, mikäli jollain alaluokalla on omia alaluokkia on niiden edessä ++ jne. Parhaiten sanaston muodon ymmärtää kuvan 12. avulla. Olennaista on huomata, että STT:n nykyisessä järjestelmässä luokkia ei ole yksiselitteisesti koodattu, esimerkiksi numerokoodeilla, kuten IPTC-luokituksessa, mikä helpottaisi tiedon analysointia koneellisesti. Luokkien numeroimattomuus tuo myös mukanaan moniselitteisyysongelman, sillä esimerkiksi sana "Olympia" esiintyy sanastossa luokan nimenä 42 kertaa, joten pelkästään yhden annetun alaluokan perusteella ei voi aina yksikäsitteisesti tietää mihin ylempään luokkaan uutinen kuuluu, eli tässä tapauksessa on kysymys olympiatason urheilusta, mutta ei voida tietää, mistä lajista puhutaan.



Kuva 12. Esimerkki luokkahierarkiasta STT:n käyttämässä luokittelusanastossa

Opetusaineisto

Monet automaattiset luokittelujärjestelmät perustuvat siihen, että ne poimivat käsin luokitellusta ns. opetusmateriaalista piirteitä, joidenka perusteella ne muodostavat kullekin luokittelusanaston luokalle omanlaisensa piirreprofiilin. Tämän jälkeen uudesta luokittelemattomasta artikkelista poimittuja piirteitä verrataan luokkien profiileihin ja sijoitetaan artikkeli parhaiten sopivaan luokkaan. Toimiakseen tämänkaltainen tilastollisuuteen perustuva järjestelmä tarvitsee kullekin luokalle riittävästi mahdollisimman hyvin kyseistä luokkaa kuvaavia artikkeleita opetusmateriaalikseen. Testatuista järjestelmistä yhtä lukuun ottamatta kaikki muut käyttävät tällaista tilastollista lähestymistapaa luokitteluun.

Opetusaineistoksi toimitettiin noin 50 000 vuosina 2004–2005 julkaistua uutista. Uutiset toimitettiin erillisinä XML-tiedostoina, joissa oli mukana yksilöllinen tunnistenumero, uutisen prioriteettiarvo, category-tagin sisällä: toimittajien antamat luokitukset uutiselle, uutisen osasto, otsikko, uutisteksti ja modified-tagin sisällä aikakoodi, joka kertoo, milloin uutista on viimeksi muutettu. Esimerkki opetusaineistona käytetystä uutisesta löytyy liitteestä B.

Testiaineisto

Testiaineistona toimitettiin tuhat uutista, jotka olivat muuten samankaltaisia, kuin opetusaineistonkin uutiset, mutta niistä oli poistettu luokittelutiedot, eli category-tagin sisältö. Vaikka testiaineistoon pyrittiin valitsemaan uutisia tasapuolisesti eri osastoilta ja näin saamaan kattavuutta, niin kaikista luokittelusanaston yli tuhannesta luokasta testimateriaalissa esiintyi vain 291 luokkaa. Ja edelleen näistäkin luokista suurin osa esiintyi vain muutaman kerran. Kaikkein yleisimpänä luokkana testiaineistossa oli politiikka 329 esiintymällä (Kuva 2.). Voidaan olettaa, että koko STT:n arkiston ja näin ollen myös 50 000 artikkelin opetusaineiston luokittelutietojen jakauma on varsin epätasaisesti jakautunutta.

Kuva 13. Kaksikymmentä yleisintä luokkaa ja niiden esiintymismäärät testiaineistossa. Plus-merkki luokan nimen edessä tarkoittaa, että se on ensimmäisen tason alaluokka (esimerkiksi puolueet on politiikan alaluokka).

Kuva 14. Kaikkien luokkien jakauma testiaineistossa pääluokkien mukaan esitettynä. Kuvaajassa olevat luvut ovat pääluokan hierarkkisesti ja sen alaisten alaluokkien esiintymien summia

Luokittelutestin vertailumenetelmät

Yleisesti hakutuloksia, indeksointia, luokittelua ym. arvioidaan menetelmillä, jotka perustuvat relevanssin käsitteeseen. Relevantti tieto voidaan ymmärtää puhtaan teknisesti hakusanojen täsmäyttämisellä dokumentteihin tai sitten ns. käyttäjärelevanssin kautta, joka perustuu käyttäjän arvioon tulosten hyödyllisyydestä.

Tässä testissä on siis tutkittu sitä, kuinka relevantteja ovat eri luokittelujärjestelmien ehdottamat luokittelutiedot.

Kuten tiedonhakua, niin myös automaattista luokittelua arvioidaan yleensä käyttäen kriteerinä saantia ja tarkkuutta (katso. Luku 3). Luokittelun tapauksessa saanti on tunnusluku, joka kuvaa luokitteluehdotuksien osumien suhdetta kaikkiin relevantteihin luokkiin, eli kuinka suurta osaa kaikista oikeista luokista automaattinen luokittelija osaa ehdottaa. Tarkkuus puolestaan kuvaa sitä, kuinka suuri osuus ehdotetuista luokista koostuu relevanteista luokista, eli ehdottaako automaattinen luokittelija vain oikeita luokkia, vai myös epärelevantteja luokkia.

Saannin ja tarkkuuden suhde on yleensä käänteinen, eli saannin paraneminen johtaa tarkkuuden heikkenemiseen ja päinvastoin (Järvelin 1995) . Esimerkiksi mitä useampia luokkia automaattinen luokittelija ehdottaa sitä parempaan saantiin se todennäköisesti pääsee, mutta luonnollisesti tarkkuuden kustannuksella. Jotta eri luokittelujärjestelmien laatua voitaisiin yhteismitallisesti vertailla saannin ja tarkkuuden avulla on eri luokittelujärjestelmien antamat ehdotuslistat katkaistu aina käsintehtyjen luokitusten listan mittaisiksi. Tällöin saanti on teoriassa yhtä suuri kuin tarkkuus ja kaksi tunnuslukua voidaan esittää yhdessä ns. saannin ja tarkkuuden yhtäsuuruuspisteessä. Käytännössä saannin ja tarkkuuden lukuarvot saattavat poiketa edelleenkin hieman toisistaan, koska joissain tapauksissa alkuperäisiä luokituksia on enemmän, kuin ehdotettuja luokituksia ja myös duplikaattiluokkien sivuuttaminen vaikuttaa hieman tuloksiin. Eroavaisuudet lukuarvoissa ovat kuitenkin pieniä.

Vertailussa mukana olevia järjestelmäntoimittajia pyydettiin toimittamaan tuloksista kaksi eri listaa. Toinen, jossa olisi jo valmiiksi valikoitu jollain kriteerillä sopiva määrä luokitteluehdotuksia uutisille ja toinen, jossa olisi listattuna relevanttiusjärjestyksessä kymmenen parhaiten soveltuvaa luokkaa. Näitä listoja vertailtiin käsintehtyihin luokituksiin se. laskettiin saanti ja tarkkuus lyhyemmille optimoiduille listoille ja käytettiin pidempiä listoja saannin ja tarkkuuden yhtäsuuruuspisteen määrittämiseen. Koska luokitustietoihin tallentuneita polkuja (esim. `Urheilu, Moottoriurheilu, Moottoripyöräily`) ei ollut mahdollista tehokkaasti hyödyntää näin lyhyessä ajassa tulkittiin kaikkia luokitussanoja toisistaan irrallisina. Tulkinta johtaa siihen, että saadut saannin ja tarkkuuden numeroarvot eivät sellaisenaan kuvaa aivan aidosti tilannetta. Kaikkien järjestelmien tuloksia tulkitaan kuitenkin samalla tavalla, joten tulkintakysymys ei vaikuta testin lopputuloksena saatavaan luokittelujärjestelmien paremmuusjärjestykseen.

Toimittajan tekemiä luokituksia ei voida pitää ainoina oikeina luokituksina, vaan päinvastoin ne ovat alttiita inhimillisille virhetekijöille. Tämän takia haluttiin tutkia, onko niiden ehdotusten, joita automaattiset luokittelijat antavat joukossa sellaisia, jotka ovat oikeita, mutta joita toimittaja ei luokitusta tehdessään ole tullut ajatelleeksi. Luokitteluehdotusten relevanttiuden tarkistus tehtiin lukemalla uutisartikkeli uudelleen ja arvioimalla sen jälkeen, mitkä automaattisten luokittelijoiden antamista ehdotuksista vaikuttavat järkeviltä. Koska uutisten uudelleen lukeminen on työlästä ja aikaa vievää, valittiin tuhannen uutisen joukosta pieni 30 uutisen otos tarkempaan tarkasteluun.

Kun luokittelutuloksia oli ensin verrattu käsin tehtyihin luokituksiin, laskettiin testiaineiston tuhannelle uutiselle niiden automaattisen luokittelun vaikeutta kuvaava tunnusluku. Tämä luku saatiin summaamalla kaikkien neljän vertailussa mukana olevan järjestelmän osumat kyseisen artikkelin luokkiin ja jakamalla se käsin

luokiteltujen luokkien lukumäärällä. Tarkistusotokseen valittiin automaattisen luokituksen vaikeuden perusteella kymmenen helpointa, kymmenen keskivaikeaa ja kymmenen vaikeimmin luokiteltavaa uutista.

Kaikkien luokittelujärjestelmien antamat ehdotukset sekä STT:n käsin luokitellut ehdotukset laitettiin listaksi, josta aina uutisen lukemisen jälkeen poistettiin epärelevantit luokitteluehdotukset. Näin saatiin testiotoksen uutisille käsin luokitusta laajempi relevanttien luokkien joukko, jota sitten verrattiin uudelleen eri järjestelmien antamiin tuloksiin.

Menetelmä ei suinkaan ole aukoton, sillä tässäkin tapauksessa yksi henkilö joutuu usein hyvin tulkinnanvaraisissa tilanteissa tekemään päätöksen siitä, onko jokin luokka relevantti vai ei. Lisäksi eri ohjelmistojen tuottamia luokittelutuloksia käsitellään yhdessä, mikä saattaa hieman vääristää tulosta.

6. Haastatteluiden ja ohjelmistovertailun tulokset

Kappaleessa 6.1 mallinnetaan STT:n nykyistä uutistoimistotoimintaa informaatiojärjestelmänä Ingwersenin ja Järvelinin yhdistetyn tiedonhankinnan ja tiedonhaun mallin mukaisesti.

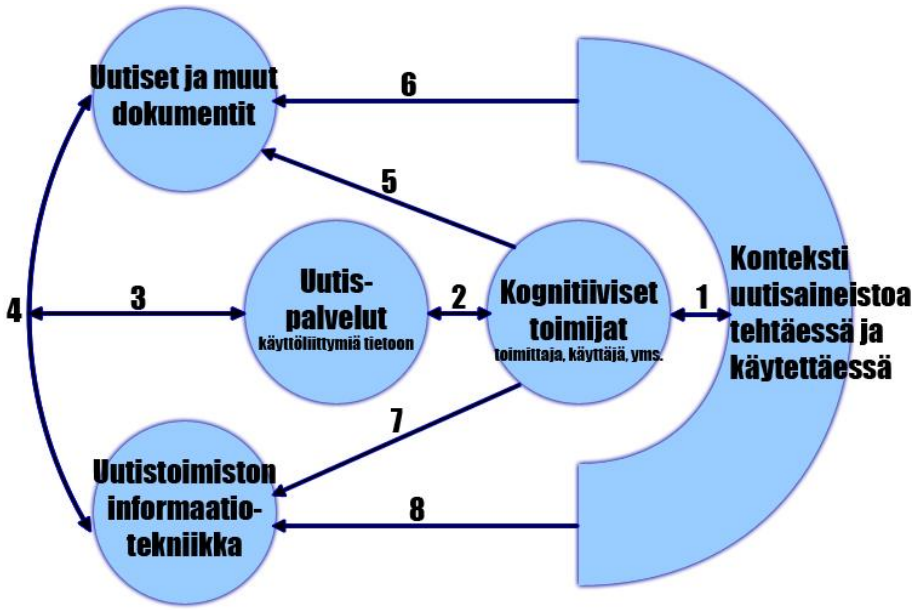
Tutkimukseen valitun näkökulman mukaisesti pyritään STT:n informaatiojärjestelmää kehittämään tunnistamalla uutisaineiston loppukäyttäjien tarpeita, sekä toimintamalleja ja kehittämällä uutispalveluita haku- ja personointiteknologioiden avulla. Asiakashaastatteluiden tulokset esitellään kappaleessa 6.2 ja haku- ja personointiteknologioiden vertailun tulokset esitellään kappaleessa 6.3.

6.1. Uutistoimisto informaatiojärjestelmänä

Informaatiotutkimuksen näkökulmasta Suomen Tietotoimisto voidaan ymmärtää informaatiojärjestelmänä. Uutistoimisto, toimittajat, käytetty informaatiotekniikka, tarjotut palvelut ja palveluita käyttävät asiakkaat ovat osana järjestelmää, jossa tietoa luodaan, tallennetaan, haetaan ja siirretään.

Pyrittäessä parantamaan STT:n palveluita on huomioitava sekä tekniset järjestelmät, jotka mahdollistavat modernit uutispalvelut, sekä palveluiden loppukäyttäjät ja toimittajat kognitiivisina toimijoina.

Tässä kappaleessa sovellan Ingwersenin ja Järvelinin yhdistettyä tiedonhankinnan ja tiedonhaun mallia (Kuva 3.) kuvaamaan STT:n informaatiojärjestelmää (Kuva 15). Malli soveltuu kuvaamaan kokonaisvaltaisesti monimutkaista informaatiojärjestelmää, jota halutaan kehittää tasapainoisesti, koska siinä huomioidaan sekä tekniset järjestelmät, kognitiiviset toimijat että toimijoiden konteksti.



Kuva 15. Yhdistetyn tiedonhankinnan ja tiedonhaun mallin (Ingwersen, Järvelin 2005) sovellus uutistoimistoon informaatiojärjestelmänä.

Seuraavaksi kuvaillaan tarkemmin STT:hen sovelletun tiedonhankinnan ja tiedonhaun mallin eri komponentit: Uutiset ja muut dokumentit, Uutistoimiston informaatiotekniikka, Uutispalvelut, käyttöliittymiä tietoon, Kognitiivinen toimijat ja Konteksti uutisia tehtäessä ja käytettäessä. Näitä komponenttikuvauksia voi verrata kappaleesta 2.4.3. löytyviin alkuperäiseen Ingwersenin ja Järvelinin mallin komponenttien kuvauksiin.

6.1.1. Uutisjutut ja muut dokumentit

Suomen Tietotoimiston informaatiojärjestelmässä käsitellään päivittäin tuhansia dokumentteja. Toimituksen rooli on valikoida, jalostaa ja välittää eri kanavia pitkin sisään tulevaa informaatiota, asiakkaiden tarpeita mahdollisimman hyvin vastaaviksi uutispalveluiksi. Suuri osa toimitukseen tulevasta informaatiosta on jo valmiiksi jossain digitaalisessa muodossa olevina dokumentteina, kuten ulkomaisten uutistoimistojen materiaali, lehdistötiedotteet ja sähköpostit mutta lisäksi toimittajat osallistuvat tiedotustilaisuuksiin ja hankkivat tietoa muista henkilölähteistä.

Dokumenteilla tarkoitetaan tässä tutkimuksessa rajatusti vain niitä dokumentteja, jotka ovat loppukäyttäjille saatavilla STT:n tarjoamien kaupallisten ja ei-kaupallisten palveluiden kautta. Informaatiojärjestelmän kokonaisvaltaisessa parantamisessa on toki huomioitava myös toimittajien tarpeet ja sitä kautta myös toimitukseen saapuvien dokumenttien ns. raakamateriaalin käsittely, sillä toimittajien työn helpottuminen heijastuu myös loppukäyttäjille palveluiden laadun paranemisena.

Pääosa STT:n asiakkaille julkaisemista dokumenteista on uutisjuttuja, eli jostain uutiskynnyksen ylittävästä tapahtumasta kirjoitettuja juttuja. Uutistoimiston tehtävänä on tuottaa nopeasti lähdemateriaalia muiden viestimien käytettäväksi. Tästä johtuen usein samasta tapahtumasta julkaistaan useita uutisjuttuja, jolloin myöhempiin versioihin on voitu lisätä tarkempia ja uudempia tietoja asiasta. Osasta uutisista tehdään myös eri julkaisukanavia varten erillisiä uutisjuttuja, kuten lyhennetty tekstiviestiversio, suoraan verkkojulkaisuun tarkoitettu uutisjuttu ja tavallinen uutispalvelun kautta lähetettävä uutisjuttu.

Tekstimuotoisten uutisjuttujen ohella STT julkaisee muitakin dokumentteja, kuten grafiikkaa, ääniuutisia, kokonaisia artikkeleita sisältäen tekstin ja niihin liittyviä grafiikoita. Tiedotepalvelun kautta STT välittää myös kaupallisia tiedotteita, jotka eivät ole journalistisen työn tuloksena syntyneitä uutisjuttuja.

6.1.2. Uutistoimiston informaatiotekniikka

Uutistoimiston käytössä olevat uutisdokumenttien käsittelyyn osallistuvat laitteistot ja ohjelmat voidaan jakaa toiminnallisuuden perusteella karkeasti tiedostonhallintajärjestelmään, toimitusjärjestelmään, ja jakelujärjestelmään. Käytännön toteutuksessa näiden tehtäväalueiden hoitamisesta saattaa vastata yksi tai useampi ohjelmisto se. esimerkiksi tiedostonhallinta on sulautettuna toimitusjärjestelmään tai kaikki on erotettu toisistaan ja lisäksi tiedostonhallinnasta huolehtii yhdessä dokumenttien tallennusjärjestelmä ja erillinen hakukone, joka indeksoi ja rikastaa tietoa, sekä toteuttaa hakuja.

Ingwersenin ja Järvelinin yhdistetyn tiedonhaun ja -hankinnan mallin mukaan ainoastaan tiedostonhallintajärjestelmä hakutoiminnallisuuksineen kuuluu IT-komponenttiin, kun taas toimitusjärjestelmän ja jakelujärjestelmän avulla toteutetaan käyttöliittymiä samaan tietovarastoon.

Tutkimusta tehtäessä ei rajoituta ajattelemaan STT:n informaatiojärjestelmän IT-komponenttia sellaisena, kuin se tutkimuksen tekohetkellä on, vaan pyrkimyksenä on tuottaa tutkimustuloksia, joita voidaan hyödyntää nykyisen järjestelmän parantamisessa. Käytössä olevat ratkaisut antavat kuitenkin suuntaviivoja siihen, mitä nimenomaan uutistoimiston IT-ratkaisut yleensä sisältävät.

Informaatiotekniikan osalta uutistoimisto ei poikkea muista informaatiojärjestelmistä. Keskeisimmässä osassa uutistoimiston IT-ratkaisuissa ovat käytetty tiedostonhallintajärjestelmä, sekä haku- ja luokitteluteknologiat.

6.1.3. Uutispalvelut, erilaisia käyttöliittymiä tietoon

Julkaisujärjestelmällä toteutetaan erilaiset asiakkaille suunnatut käyttöliittymät tietovarantoon. Toimitusjärjestelmä, jolla uusia dokumentteja luodaan on puolestaan toimittajien oma käyttöliittymä dokumenttivarastoon. Toimitusjärjestelmän ja julkaisujärjestelmän ei tarvitse välttämättä olla teknisesti toisistaan erillisiä, vaan loogisina toimintakokonaisuuksina ne voidaan toteuttaa hyvinkin monenlaisilla järjestelmäarkkitehtuureilla.

STT:n tapauksessa kaikki tarjotut kaupalliset palvelut hyödyntävät samaa tietovarantoa, mutta tarjoavat erilaiset käyttöliittymät loppukäyttäjille. STT:n toiminnan ydin on uutispalvelu, jota toimitetaan vuorokauden vuoden jokaisena päivänä.

Palveluiden ja sitä kautta myös käyttöliittymien eroina ovat mm. se, mihin osaan informaatiosta niiden kautta on mahdollista päästä käsiksi, missä formaatissa informaatio toimitetaan, kuinka paljon käyttäjällä on kontrollia (ns. push- ja pull-palvelut) ja millä fyysisellä välineellä palvelua käytetään.

Seuraavassa on lyhyet kuvaukset tutkimuksen kannalta oleellisista palveluista (Taulukko 6), joita STT tutkimuksen tekohetkellä tarjosi asiakkailleen, suluissa on palvelun kaupallinen nimi.

Taulukko 6. Kuvaukset tutkimuksen kannalta oleellisista STT:n tutkimuksen tekoaikana tarjoamista palveluista.

Perus uutispalvelu	Kaikki uusimmat uutiset uutisosastojen (esim. kotimaa ja urheilu) mukaisesti jaoteltuina.
Tekstiviestiuutispalvelu	Lyhennetyt versiot eri osastojen tärkeimmistä uutisista tekstiviestinä.
Toimialaseuranta	Asiakaskohtaisesti räätälöity uutispalvelu, jossa tietyn hakuehdon täyttävät uutiset toimitetaan asiakkaalle reaaliaikaisesti.
Tiedotepalvelu	Tiedottajille maksullinen palvelu, välittää esimerkiksi kaupallisia tiedotteita toimittajille.
Arkistopalvelu	Arkisto STT:n julkaisemista uutisista
Listat ja kalenterit	Listat ja kalenterit ovat median suunnittelutyökaluja, joista näkee tulevia uutistapahtumia.

Perus uutispalvelu (STT Uutispalvelu)

Perus uutispalvelun tilaajalla on jatkuvasti pääsy uusimpiin enintään muutaman viikon vanhoihin STT:n välittämiin uutisiin. Palvelun tilaaja voi tilata halutessaan kaikki tai vain tiettyjen osastojen, kuten kotimaa, urheilu, kulttuuri yms. uutiset. Media-asiakkaille uutiset toimitetaan yleensä suoraan toimitusjärjestelmään (push-palvelu) näiden vaatimassa teknisessä muodossa, jolloin käyttöliittymä rakennetaan osaksi asiakkaan toimitusjärjestelmää. Tutkimuksen kohderyhmänä olevat median ulkopuoliset asiakkaat käyttävät kaikki kuitenkin perus uutispalvelua omalla salasanallaan STT:n verkkosivuille rakennetulla käyttöliittymällä, jolloin kyseessä on push-palvelu.

Uutisia voidaan toimittaa osastoittain (kotimaan uutiset, ulkomaan uutiset, talous-, politiikka-, urheilu-, kulttuuri- ja viihdeuutiset sekä urheilutulokset), asiasanojen perusteella (esimerkiksi vaalit, media, metsäteollisuus) alueluokituksella (esimerkiksi Suomi, Etelä-Pohjanmaa, EU, Yhdysvallat) tai vapaasanahaun perusteella. Verkkokäyttöliittymän kautta tilaajalla on pääsy kahta viikkoa uudempaan uutismateriaaliin valituilta osastoilta.

Tekstiviestiuutispalvelu (STT Mobiiliuutiset)

Tekstiviestiuutispalvelun tilaajat saavat tilaamiensa osastojen pääuutisista lyhennetyt versiot tekstiviestillä suoraan matkapuhelimeensa. STT:n toimitus valitsee tekstiviestinä välitettävät uutiset niiden tärkeyden mukaan ja tekee niistä lyhennetyt versiot tekstiviestivälitystä varten.

Uutiset toimitetaan joko tekstiviesteinä (push-palvelu) tai internetistä noudettavina wap-uutisina (pull-palvelu). Palvelua tilatessaan asiakas määrittelee, minkä kategorioiden uutisia hän haluaa saada tekstiviestinä. Tyypillisesti palvelu koostuu pääuutisten kategoriasta ja sitä täydentävistä erityiskategorioista, kuten ”Pörssi” ja ”Formula 1”. Kaikkiaan valittavia uutiskategorioita on yksitoista. Pääuutisissa ovat päivän tärkeimmät ja kiinnostavimmat uutiset aihepiiristä riippumatta. Keskimäärin pääuutisia on päivittäin 1 - 2.

Toimialaseuranta (STT Uutisvahti)

Toimialaseuranta on nimi nykyiselle asiakaskohtaisesti räätälöidylle uutispalvelulle, jossa jonkun tietyn hakuehdon täyttävät uutiset valikoidaan automaattisesti ja toimitetaan asiakkaan haluamassa muodossa esimerkiksi julkaistavaksi intranetissä tai verkkosivuilla.

Palvelussa asiakkaalle määritellään hakusanoin seurattavat aiheet. Ne voivat olla esimerkiksi yritysnimiä tai toimialaa kuvaavia sanoja. Kun STT välittää valitut kriteerit täyttävän uutisen, asiakas saa siitä oman version haluamallaan tavalla automaattisesti (push-palvelu).

Toimitusformaattina voi olla esimerkiksi uutisen otsikko tekstiviestinä ja koko uutinen sähköpostiin lähetettynä.

Tiedotepalvelu (STT Tiedotepalvelu)

Tiedotepalvelu on tiedottajille maksullinen kanava välittää esimerkiksi kaupallisia tiedotteita, jotka eivät muuten välttämättä ylittäisi uutiskynnystä. Tiedotepalvelun kautta lähetetyt tiedotteet välitetään rekisteröityneille toimittajille ilmaiseksi. Tiedotepalvelu ei ole käyttöliittymä, jonka kautta pääsisi käsiksi STT:n toimituksellisesti tuottamaan informaatioon, mutta erityisesti tiedottajat käyttävät sitä

yhdessä muiden STT:n palveluiden kanssa, joten sitä ei voida tutkimuksessa sivuuttaa.

Tiedoteportaali on suunniteltu palvelemaan STT:n media-asiakkaita eli lehdistöä sekä televisio- ja radiokanavia siten, että he voivat vastaanottaa valmiiksi luokiteltua tiedoteaineistoa yrityksiltä, järjestöiltä ja julkishallinnolta.

Tiedotepalvelun kautta on mahdollista lähettää tiedotteita myös STT:n uutispalvelun mukana suoraan median toimitusjärjestelmiin. Uutispalvelun kautta välitettyjen tiedotteiden tulee olla uutis- tai informaatioarvoisia.

Arkistopalvelu (STT Uutisarkisto)

Arkistopalvelu on STT:n verkkosivuille rakennettu käyttöliittymä, jonka kautta palvelun tilaajilla on mahdollista tehdä hakuja vanhempaan uutismateriaaliin (pull-palvelu), joka on joskus julkaistu perus uutispalvelun kautta, mutta ei ole enää saatavilla sieltä. STT:n uutisarkistossa on tallennettuna tietotoimiston uutismateriaali vuodesta 1991. Selailun helpottamiseksi arkiston materiaali on jaettu osastoittain. Uutisarkisto on verkkopalvelu, jonka käyttö vaatii erilliset tunnukset.

Listat ja kalenterit (STT Listat ja Kalenterit)

STT:n päivälista on pääasiassa media-asiakkaille suunnattu suunnittelutyökalu, jolla STT informoi uutisarvoisista aiheista. Se kertoo päivittäin tärkeimmät toimituksia kiinnostavat tapahtumat, kuten esimerkiksi tiedotustilaisuudet. STT ei itse aktiivisesti etsi listoille tapahtumia, vaan yhteisöt, yritykset ja järjestöt ilmoittavat itse STT:lle tiedotus- ja muista tilaisuuksistaan. STT:n toimitus valikoi tapahtumat listalle journalistisin perustein. Luonnollisesti yllättävät uutisaiheet, kuten onnettomuudet tai vastaavat eivät ole listalla. Suuri osa päivän uutisaiheista aina urheilutapahtumista EU-kokouksiin on kuitenkin mahdollista tietää etukäteen.

Tapahtuma- urheilu ja ulkomaankalentereihin on poimittu perustiedot tapahtumista kuten kongresseista, messuista, ensi-illoista, festivaaleista ja muista yleisö- ja uutistapahtumista.

Median ulkopuoliset asiakkaat käyttävät listojen ja kalentereiden salasanalla aukeavia verkkoversioita. Kalentereista ja listoista toimitetaan myös sähköpostiversiot ja media-asiakkaat voivat saada tiedot suoraan toimitusjärjestelmiinsä.

6.1.4. Kognitiiviset toimijat

Pääasialliset uutistoimistoon liittyvät kognitiiviset toimijat ovat:

Uutisia, uutisaiheita ja lähdemateriaalia tuottavat tahot, kuten muiden uutistoimistojen työntekijät, kirjeenvaihtajat ja tiedottajat ja uutisvinkkejä antava yleisö

Uutisia tekevät ja muokkaavat toimittajat, joidenka tehtävänä on myös päättää uutisten julkaisusta ja prioriteetista

Uutistoimiston IT-ratkaisuja ja palveluita suunnittelevat ja toteuttavat henkilöt

Uutisaineiston loppukäyttäjät

Sama henkilö saattaa olla kognitiivisena toimijana eri rooleissa. On esimerkiksi hyvin tavallista, että toimittaja oman työnsä puitteissa on informaatiojärjestelmässä

sekä tiedon tuottajana, että käyttäjänä. Kognitiivisten toimijoiden joukko voi helposti kasvaa hyvinkin laajaksi.

Tässä työssä STT:n informaatiojärjestelmää on tutkittu uutispalveluita työtehtävissään käyttävän henkilön, eli uutisaineiston loppukäyttäjän näkökulmasta. Loppukäyttäjät ovat tiedonhakijoita eli kognitiivisia toimijoita ja siten osa informaatiojärjestelmää.

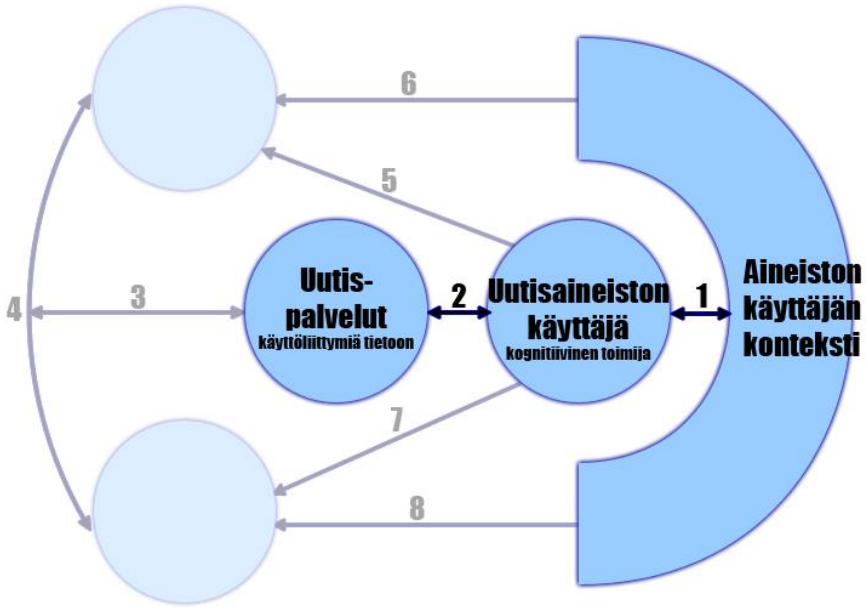
6.1.5. Konteksti uutisia tehtäessä ja käytettäessä

Tämän tutkimuksen kannalta merkittävin konteksti on tiedonhankkijan sosiaalinen, organisatorinen ja kulttuurillinen konteksti. Tiedonhankkija käsitetään tavoitteellisesti toimivaksi yksilöksi, joka etsii ja hyödyntää tietoa sen merkityksen nojalla niistä lähteistä ja kanavista, joiden hän uskoo parhaiten tyydyttävän tiedontarpeitaan (Savolainen 2000). Nimenomaan huomio tiedon merkityksestä edellyttää tiedonhankkijan kontekstin huomioimista, sillä sama informaatiosisältö saattaa saada eri tiedonhankkijoiden tulkitsemana hyvinkin erilaisia merkityksiä ja jopa yksi ja sama tiedonhankkija voi tulkita tilanteesta riippuen saman informaation eri tavoin.

Tiedonhankkijan konteksti vaikuttaa välitettävän informaation tulkinnan lisäksi myös siihen, kuinka tyytyväinen käyttäjä on palveluun. Esimerkiksi palveluiden hinnan voidaan ajatella olevan kontekstuaalinen tekijä, mikäli loppukäyttäjä itse on vastuussa uutispalvelun ostamisesta tai ainakin tietoinen palvelun hinnasta. Palvelua käyttävän asiakkaan reaktiot esimerkiksi hinnoitteluun puolestaan riippuvat kyseisen käyttäjän asemasta omassa organisaatiossaan, eli hänen organisatorisesta kontekstista.

Myös toimittajien konteksti uutisjuttuja tehtäessä vaikuttaa merkittävästi koko tutkittavan uutistoimiston informaatiojärjestelmän toimintaan. Esimerkiksi toimituksen menetelmät ja käytännöt vaikuttavat siihen, minkälainen uutinen ylittää uutiskynnyksen.

6.2. Haastattelututkimuksen tulokset



Kuva 16. Haastattelututkimuksessa keskityttiin uutisaineiston käyttäjään kognitiivisena toimijana ja pyrittiin selvittämään uutispalveluiden käyttöä, sekä niitä kontekstuaalisia tekijöitä, jotka vaikuttavat uutisaineiston käyttäjään.

Haastatteluaineistot ovat erittäin rikas idealähde siitä, minkälaiset palvelut ovat käyttäjien kannalta hyödyllisimpiä. Tässä kappaleessa pyrin esittämään mahdollisimman elävästi haastatteluiden keskeisimmän annin. Esityksen tukena käytän suoria lainauksia haastateltavilta.

Haastatteluiden perusteella voidaan erottaa neljä selvästi toisistaan poikkeavaa käyttötarkoitusta STT:n palveluille; uutisseuranta, mediaviestintä, tiedonhaku ja julkaisutoiminta. Nämä käyttötarkoitukset toimivat yläkäsitteinä erilaisille toimintatavoille ja joskus yksityiskohtaisemmallekin käyttötarkoitusten jaottelulle. Esimerkiksi uutisseurantaa STT:n perus uutispalvelun avulla saatetaan tehdä kollektiivisesti (toimintatapa) tai sen takia, että voidaan reagoida nopeasti tarvittaessa (yksityiskohtainen käyttötarkoitus). Samaa tarkoitusta varten saatetaan käyttää rinnakkain useita eri palveluita (katso Taulukko 6 STT:n tarjoamista palveluista). Esimerkiksi mediaviestintään käytetään niin perus uutispalvelua, tiedotepalvelua kuin listoja ja kalenteriakin. Eri käyttötarkoitukset ja niihin liittyvät toimintarutiinit paljastavat myös niitä asioita, joissa palveluita voisi kehittää nykyisestään. Seuraavassa käsitellään hieman yksityiskohtaisemmin kutakin näistä.

6.2.1. Uutisseuranta

Uutisseurannalla tarkoitetaan sekä STT:n uutispalvelun, että muiden uutislähteiden kautta tapahtuvan päivittäisten uutisten seurantaa. Käyttäjillä on oman työnsä puolesta ja henkilökohtaisestikin tiettyjä kiinnostuksenkohteita, joihin liittyviä artikkeleita päivän uutisvirrasta haetaan. Erotuksena mediaseurannasta uutisseurannalla tarkoitetaan vapaamuotoisempaa uutisten selailua, joka riippuu aina päivän muista työkiireistä, eikä ole tiukasti ennalta määrättyihin aiheisiin sidottua. Mediaseuranta on eri uutislähteiden systemaattista läpikäyntiä, jossa pyritään

löytämään kaikki jostain aiheesta kirjoitetut uutiset. Markkinoilla on mediaseurantaan erikoistuneita yrityksiä, joidenka palveluita myös jotkut haastateltavat käyttivät STT:n palveluiden ohella.

Uutisia seurataan jatkuvasti, mutta usein haastateltavat eivät osaa yksiselitteisesti perustella, minkä takia seuraavat uutisia. Uutisseurannan koetaan olevan hyödyllistä ja jopa välttämätöntä. Haastatellut puhuvat kärryillä pysymisestä ja ammatillisesta yleissivistyksestä. Yleensä he perustelevat uutisseurantaansa sillä, että heidän on tiedettävä erinäisiä asioita maailman tapahtumista. Se mitä tällä tiedolla tehdään, on kuitenkin hankalasti sanoiksi puettavissa. Uutisseuranta vastaa tyypillisesti orientoivaan tiedontarpeeseen.

Suomalaisen yhteiskunnan arvot vaikuttavat taustalla vahvasti siihen, miksi uutisten seuraamista pidetään tärkeänä. Sen takia monet työssään STT:n uutispalvelua käyttävät seuraavat palvelua myös omasta mielenkiinnosta, vaikkei se olennaisesti liittyisikään työhön.

> *"viestinnän puolella pitäs yrittää pystyä seuraamaan, et mitä maailmassa ja Suomessa tapahtuu [...] mutta tietysti ihan omasta mielenkiinnostakin se on hyvä"* -Haastateltava 8.

Useiden haastateltujen kommentteja lukemalla hahmottui kaksi pääasiallista työtehtäviin liittyvää syytä uutisseurantaan, jotka ovat: varautuminen ja reagointi. Eräs usein toistuva toimintatapa oli kollektiivinen uutisseuranta, jolloin uutisia seurataan toisten ihmisten puolesta. Uutispalvelua käyttävä henkilö ei siis aina itse välttämättä reagoi uutiseen tai se ei ole merkittävä edes sen kannalta, että hän voisi varautua tulevaisuuden työtehtäviin, mutta hän saattaa välittää tiedon eteenpäin henkilölle, joka reagoi tai varautuu. Tällaista kollektiivista uutisseurantaa ja tiedon edelleenvälitystä tapahtuu kaikenaikaa ja usein se on tärkeä osa organisaatioiden toimintaa.

> *"jos mä huomaan jotain [...] merkittävämpää, niin kyl mä yleensä sitten informoin välittömästi [...] ja laitan ihmisille viestiä"* -Haastateltava 4.

> *"koska konsultit on aika paljon liikkeellä, ni sit itekkin tulee käytyä niitä [uutisia] läpi ja sit mä heitän, [...] et hei et ootsä huomannu, et sun asiakkaast oli tämmönen juttu"* - Haastateltava 2.

Varautuminen

Suurin osa uutisoinnista ei johda mihinkään välittömiin toimiin uutispalveluiden käyttäjien keskuudessa tai heidän organisaatioissaan. Tällöin uutisia seurataan, jotta voidaan varautua mahdollisiin yhteydenottoihin tai muihin uutisten aiheuttamiin työtehtäviin, jotka syntyvät yleensä viipeellä, kun joku muukin on nähnyt saman uutisen.

> *"Se on oikeestaan [...] valmiuden luomista. Et meiän hommas on hyvin tärkeetä, et tietää, missä mennään ja mitä tapahtuu, koska se heijastuu näihin meiän toimeksiantoihin."* - Haastateltava 1.

> *"Mä esimerkiksi huomaan, ett nyt on tästä aiheesta joku puhunu tämmöstä ja tämmöstä [...] meillä täytyy olla semmonen varautuminen, että ehkä joku ottaa vaikka yhteyttä, kysyy meidän kantaa tähän asiaan."* -Haastateltava 4.

Reagointi

Pieni osa uutisista on sellaisia, että uutispalvelun käyttäjä reagoi niihin välittömästi. Useimmiten reagointi liittyy siihen, että media on yhteiskunnallisen keskustelun kenttä ja uutispalvelun käyttäjä tai hänen organisaationsa haluaa osallistua keskusteluun antamalla vastineensa johonkin uutiseen.

> *"Mulle [tulee] kännykkään tieto, eli mä nään reaaliajassa mitä STT:ltä on lähteny. Että kun joskus ihmetellään miten me niin nopeasti reagoidaan johonkin asiaan niin se johtuu siitä."* -Haastateltava 7.

> *"Jos nähdään siinä jotain ristiriitaa, että meistä välittyy sellainen kuva suurelle yleisölle, joka ei meidän mielestä vastaa sitä todellisuutta, niin silloinhan meidän täytyy ryhtyä toimenpiteisiin ja miettiä miten me pystyttäis vaikuttamaan siihen kuvaan."* -Haastateltava 6.

6.2.2. Mediaviestintä

Mediaviestinnällä tarkoitetaan tässä organisaation ulospäin suuntautuvan viestinnän koko ketjun seurantaa ja hallinnointia. Viestinnässä yhtenä kanavana tai keinona on lehdistötiedottaminen, joka käsittää yleensä tiedotteiden kirjoittamisen, lähettämisen ja läpimenon seurannan. Tiedottamisessa käytetään apuna mm. STT:n tiedotepalvelua tiedotteiden välittämiseen ja uutispalvelua niiden läpimenon seurantaan STT:n osalta.

Tiedottajien kannalta STT:n rooli uutislähteenä ja tiedotteiden välityspalveluna on hieman kaksijakoinen. Se, että STT tekee organisaation lähettämän tiedotteen perusteella uutisen tai se, että tiedote välitetään sellaisenaan muille medioille STT:n tiedotepalvelun kautta palvelevat usein tiedottajan kannalta samaa päämäärää.

Joskus viesti välittyy tiedotepalvelun kautta muille medioille:

> *"jos tiedotteessa on ihan oikea uutinen niin STT [tiedotepalvelu] on aika tehokas palvelu, että olen huomannut että ennekuin STT on itse tehnyt mitään, niin se on jo YLE 24:llä"* -Haastateltava 7.

Joskus ylittyy STT:n oma uutiskynnys, kuten seuraava kommentti kuvaa:

> *"tiedotteet jotka me lähetetään STT:n toimitukselle [...] tapaa olla sen kaltaisia, että ne ylittää uutiskynnyksen ja tulevat osaksi sitä uutisfeediä"* -Haastateltava 6.

Mediaviestinnän suunnittelu ja kehittäminen

Erityisesti STT:n lista- ja kalenteripalveluita käytetään myös viestinnän suunnitteluun ja kehittämiseen. Sitä kautta seurataan, median liikkeitä ja myös muiden organisaatioiden käyttäytymistä ja pyritään kehittämään omaa viestintää paremmaksi. Listoja voidaan hyödyntää mm. tilaisuuksien ajoittamiseen, sekä oman organisaation ja muiden organisaatioiden mediaviestinnän vertailuun, kuten eräs haastateltava kommenteissaan osuvasti esittää:

> *"päivälistalta [...] pystyy myös hyvin näkee, et jos on hirveen täys päivä, ni voi arvata, että meidän tilaisuuteen ei välttämättä tuu niin montaa toimittajaa"* -Haastateltava 8.

"joskus katon myöskin [...] minkälaisista aiheista muut organisaatiot järjestää tiedotustilaisuuksia [...] tätähän [listaa] voi käyttää niinku benchmarkkaukseenkin" -Haastateltava 8.

Viestinnästä vastaavat henkilöt seuraavat aktiivisesti uutisointia myös tavoitteenaan ymmärtää paremmin median toimintaa ja sitä kautta saada valmiuksia omaan työhönsä.

"jotain aavistusta on siitä, et minkälaiset jutut menee ylipäänsä läpi, et mitä kannattaa tiedottaa [...] ku täält lähtee tiedote, ni kyl aika nopeesti tietää, et mis se menee läpi ja meneekö se läpi ylipäätään." -Haastateltava 2.

Tiedottaminen

Mediaviestinnässä keskeisessä asemassa on lehdistötiedotteiden kirjoittaminen ja lähettäminen. STT:n tiedotepalvelua käytetään yhtenä kanavana tiedotteiden levittämiseen, mutta sen rinnalla useimmat viestinnästä vastaavat lähestyvät eri lehtien ja muiden medioiden toimituksia suoraan mm. sähköpostitse.

"Sitten ku me pistetään tiedote, ni sen lisäks, että me pistetään se STT:lle, ni jokainen tiedote sitten erillisen harkinnan mukaan menee tietyillä jakeluilla sähköpostilla sitten toimituksille suoraan, eri lehtien tiedotusvälineiden toimitukseen." -Haastateltava 4.

Lehdistötiedottamisella organisaatiot pyrkivät siihen, että STT tai joku muu kirjoittaa aiheesta jutun ja organisaation tunnettavuus kasvaa. Tähän tavoitteeseen päästäkseen tiedotuksesta vastaavat henkilöt pyrkivät mahdollisimman hyvin palvelemaan toimittajia. He mm. antavat haastatteluita ja muuta ennakkoinformaatiota, kuten kaksi alla olevaa kommenttia osoittaa:

"siinä toimi STT:n kanssa asiat hienosti, eli aina kun oli uutta kerrottavaa, niin soitin STT:lle ja kerroin, että nyt semmosta ja tämmöstä" -Haastateltava 6.

"parempi niinku, että löytää ne kiinnostuneet toimittajat ja toimitukset ja [...] antaa ennakkoon sitä informaatiota ja tavallaan sitä kautta myös varmistaa, että sieltä tulee se juttu" -Haastateltava 8.

Viestien läpimenon ja julkisuuskuvan seuranta

Julkisuus on organisaatioille keino saavuttaa päämääriään ja tiedottajien vastuulla on positiivisen julkisuuden maksimointi. Aktiivisen tiedottamisen ohella organisaatioissa seurataan tiiviisti, miten viestit välittyvät medialle ja mitä loppujen lopuksi kirjoitetaan. Silloin kun STT:n toimitusta on lähestytty lehdistötiedotteella, seurataan tiedotteen mahdollista läpimenoa STT:n uutispalveluiden kautta. Organisaatioissa ollaan kiinnostuneita siitä mitä heistä julkisuudessa kirjoitellaan ja tässä suhteessa STT:n uutisointia seurataan vastaavalla tavalla kuin muitakin medioita. Useiden haastateltujen organisaatiot ostavat lisäksi palveluita erityisesti mediaseurantaan erikoistuneilta yrityksiltä. Seuraavat kommentit kuvaavat tyypillisiä viestien läpimenon ja julkisuuskuvan seurantaan liittyviä toimintoja.

"Meillä täytyy olla tarkka tieto siitä mitä meistä päivittäin kirjoitetaan" -Haastateltava 6.

"mennään STT:n sivulle [...] ja seurataan, et tuleeks se [tiedote] sinne ja jos se tulee, ni miten se on kirjotettu, et onks se menny sellasenaan läpi vai onko sitä jollakin tapaa muokattu" -Haastateltava 2.

"julkisuuskuvasta huolehtiminen on yks keskeinen tehtävä, niin paljon kun siitä pystyy tänä päivänä huolehtimaan [...] mediaseuranta on ihan sitä mitä teen päivittäin ja siinä mulla on yhteistyökumppanina [mainitsee kaksi mediaseurantaan erikoistunutta yritystä]" -Haastateltava 7.

6.2.3. Tiedonhaku

Tiedonhaku liittyy useimmiten johonkin laajempaan työtehtävään, kuten artikkelin kirjoittamiseen, asiakashankintaan tai viestintään, mutta se voi esiintyä myös omana tehtävänään, jolloin on nimenomaan vain tarkoitus selvittää joku asia ja kerätä tietoa. Tiedonhaussa STT:n rooli on huomattavasti pienempi, kuin viestinnässä ja uutisseurannassa. STT on omimmillaan tuoreen uutistiedon nopeassa välityksessä, kun taas historiallisen arkistotiedon saamiseksi on olemassa lukuisia muitakin lähteitä. Yhtenä syynä STT:n arkiston vähäiseen käyttöön on sen hitaus ja huono käytettävyys.

Vanhat kirjoitukset viestinnän apuna:

"monta kertaa voi olla, et joku asia nousee niinku uudelleen pinnalle, ni sillon ois kiva niinku kattoo, et mitä täst aiheest niinku viimekerralla kirjotettiinkaan" -Haastateltava 8.

Taustatietoa potentiaalisista asiakkaista:

"me halutaan ymmärtää sitä asiakkaan bisnestä, et sillon me valitaan sinne tietyt mediat [...] sit vaan sieltä uutisarkistosta kaivetaan esimerkiks vuosi taaksepäin, et mitä on kirjotettu ja mitkä niist on oleellisia juttuja" - Haastateltava 2.

Tiedonhakua toimeksiannosta:

"Hän soittaa meille ja sanoo, että hommaa mulle jotain perustietoa, et mikä on vammasten asema tai mitä uutta siinä on tapahtumassa lainsäädännössä tai talouspuolella viimeaikoina? Sen jälkeen me tehdään se, eli me ruvetaan kerään tietoa eri lähteistä. Aikataulut on yleensä tiukkoja, max vuorokausi aikaa tehä, usein vähemmänkin. Sit me kerätään semmonen paketti ja viedään se hänelle." -Haastateltava 1.

6.2.4. Julkaisutoiminta

Julkaisutoiminnalla tarkoitetaan tässä yhteydessä sellaisia tilanteita, joissa STT:n materiaalia julkaistaan suoraan tai sitä käytetään tausta-aineistona jonkun julkaistavaksi tarkoitetun artikkelin kirjoittamisessa. Media-asiakkaille julkaisutoiminta on varmasti huomattavasti merkittävämpi STT:n palveluiden käyttötarkoitus, kuin median ulkopuolisille asiakkaille. Vaikka nyt haastateltiinkin median ulkopuolisia asiakkaita, niin silti esille nousi pienimuotoinen julkaisutoiminta, kuten asiakaslehdissä ja verkkosivuilla tapahtuva julkaisu.

Taustamateriaalina julkaisuja kirjoitettaessa:

"Tässä [julkaisussa] ei nyt suoraan ole STT:n aineistoa hyödynnetty, mutta tietysti aina tausta-aineistona tiedonhankinnassa STT on tärkeä" -Haastateltava 3.

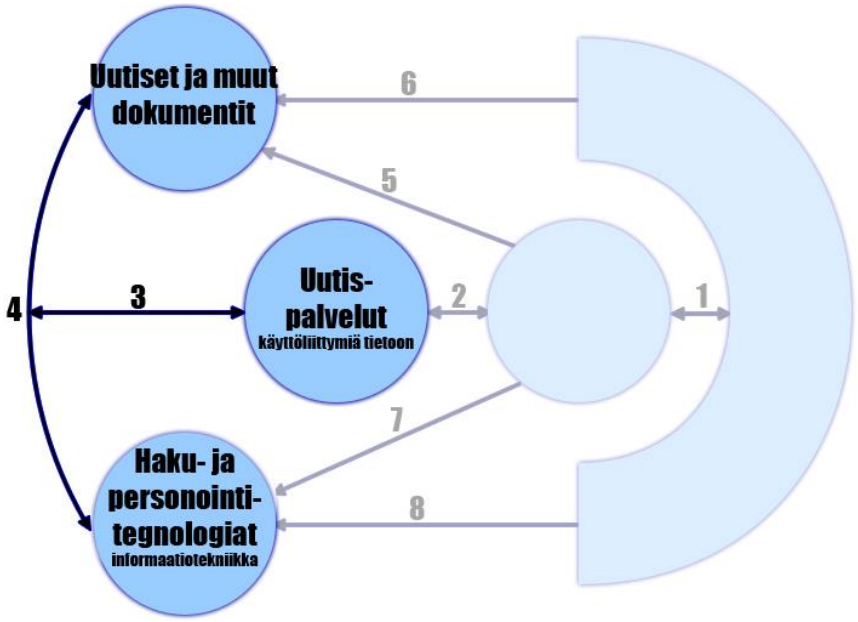
Satunnaista verkko- tai lehtijulkaisua

"[meidän] lehdessä ei oo ollut, mutta verkkopalveluissa, nytkin taitaa olla tää Suvi-Anne Siimeksestä, kun kukaan ei ehtiny paikalle sinne tilaisuuteen, niin käytettiin siinä STT:n juttuja pohjana" -Haastateltava 7.

Sisältöä julkaisuihin:

"Joo, eli he etsii siis tuota tämän Holkerin nimellä tai Kosovon perusteella uutisia ja rakentavat ja täydentävät sitä tekstiä sitten jos tarvitaan esim. taitollisista syistä tai sitten asian tärkeyden vuoksi pitää tehdä vähän laajempi juttu." -Haastateltava 3.

6.3. Ohjelmistovertailun tulokset



Kuva 17. Ohjelmistovertailun ja kirjallisuuden selvitettiin mahdollisia haku- ja personointiteknologioita, joilla voitaisiin parantaa uutispalveluiden laatua helpottamalla uutisten ja muiden dokumenttien löytymistä.

Ohjelmistovertailun ensimmäisessä vaiheessa kartoitettiin eri järjestelmien ominaisuuksia. Toisessa vaiheessa testattiin yhden ominaisuuden, automaattisen luokittelun, toimivuutta eri ohjelmistoissa. Luokittelutestin avulla vertailuun valitut ohjelmistot voidaan asettaa paremmuusjärjestykseen nimenomaan luokitteluominaisuuden toimivuuden suhteen. Tutkimuksen myöhemmässä vaiheessa tehtävät kehitysehdotukset pohjautuvat ohjelmistovertailun ensimmäisessä vaiheessa kartoitettuihin ominaisuuksiin. Luokittelutestin tuloksia ei kehitysehdotuksia tehtäessä hyödynnetä.

Ominaisuuskartoitus

Ominaisuuskartoituksessa käytiin systemaattisesti läpi ohjelmistovertailussa mukana olleiden järjestelmien ominaisuuksia. Ominaisuuksien kirjo on hyvin laaja, ja eri järjestelmissä on samantyyppiset ominaisuudet esitetty hieman eri tavalla. Lisäksi kaikki ominaisuudet eivät kuulu vakiona järjestelmiin, vaan niitä saattaa joutua lisensoimaan erikseen. Vertailua varten on yhdenmukaistettu sellaisten ominaisuuksien kuvauksia, jotka jollain tavalla vastaavat ainakin yhteen tai useampaan STT:n tarpeeseen (Taulukko 4). Tässä vaiheessa on huomioitava, että useat kartoituksen ulkopuolelle jääneet ominaisuudet varmasti ovat STT:n kannalta hyödyllisiä, mutta ne vastaavat muihin tarpeisiin, jotka eivät ole tulleet esille luokittelujärjestelmään liittyvissä keskusteluissa. Esimerkiksi joistain vertailun ohjelmistoista löytyvää web-crawleria voitaisiin hyödyntää tekijänoikeuksien valvonnassa.

Taulukko 7. Vertailtujen ohjelmistojen ominaisuuksista yhdenmukaistetut kuvaukset.

Ominaisuus 1: Nimien automaattinen merkitseminen	Järjestelmä tunnistaa ja merkitsee automaattisesti tekstissä esiintyvät yritysnimet, paikannimet ja henkilönimet ja palauttaa ne perusmuotoon ja tallentaa ne tekstin metatietoihin.
Ominaisuus 2: Automaattinen luokittelu	Automaattinen luokittelija määrittelee tekstitiedostoille täysin automaattisesti ja riittävän luotettavasti yhden tai useampia ennalta määritellyn asiasanaston mukaisia luokkia. Automaattisessa luokittelussa pitää etsiä mahdollisimman hyvä saannin ja tarkkuuden kompromissi, eli luokittelutietoihin ei saisi tulla paljoa ylimääräisiä luokkia, mutta myöskään oikeita luokkia ei saisi jäädä puuttumaan. Toimituksen omaa materiaalia luokiteltaessa automaattinen luokittelija ehdottaa, mihin kaikkiin mahdollisiin luokkiin teksti kuuluu, jonka jälkeen toimittaja valitsee mielestään oikeat vaihtoehdot. Puoliautomaattisessa luokittelussa saannin pitää olla hyvä, eli ei haittaa niin paljoa, vaikka kone ehdottaisi muutamaa vääräkin luokkaa, kunhan se ehdottaa myös kaikkia oikeita.

Ominaisuus 3: Luokittelusanaston muokattavuus	Kieli ja käytetty termistö muovautuvat ajan myötä, mikä aiheuttaa sen, että käytettyä luokittelusanastoakin pitää aika-ajoin päivittää. Usein uudet termit voivat olla sisällön kuvauksen kannalta relevantteja jo pitkään ennen kuin asiasanastoa päivitetään vastaamaan muutoksia. Esimerkiksi termi ”curling” tuskin löytyy vieläkään STT:n käyttämästä luokittelusanastosta. Luokittelusanaston muokattavuuteen liittyviä toiminnallisuuksia ovat esimerkiksi sanaston hallintaan tarkoitettu graafinen käyttöliittymä, arkistoaineiston uudelleenindeksointi uuden sanaston mukaisesti ilman järjestelmän käyttökatkosta, luokittelusanaston muutosehdotusten automaattinen tuottaminen aineistosta ja luokittelusanaston kattavuuden tilastollinen seuranta.
Ominaisuus 4: Hakuominaisuudet	Hakuominaisuudet on laaja yleiskäsite monille tiedonhakuun liittyville toiminnallisuuksille, kuten asiasanahaulle, kehittyneemmälle niin sanotulle käsitehaulle, esimerkkidokumentin kanssa semanttisesti samankaltaisten dokumenttien haulle, hakusyötteiden oikoluvulle, monikieliselle haulle jne. Tässä tutkimuksessa olen käsitellyt näitä kaikkia yhdessä.
Ominaisuus 5: Sisällön eriyttäminen	Sisällön eriyttämisellä tai personoinnilla tarkoitetaan, että eri käyttäjille voidaan automaattisesti ohjata eri sisältöä ennalta määritellyn profiilin perusteella. Sisällön eriyttämistä voidaan STT:n tapauksessa käyttää ohjaamaan sisään tulevaa uutisvirtaa eri alan toimittajille ja toisaalta erilaisten asiakkaille suunnattujen personoitujen uutispalveluiden tuottamiseen. Profilointiin liittyviä teknologioita ovat mm: käyttäjän itse täyttämä profiili, ohjelmiston määrittelemä lukutottumusprofiili, sosiaalinen muiden käyttäjien suositusten perusteella muokattava profiili ja adaptiivinen ajan myötä automaattisesti muovautuva profiili. Muita eriyttämiseen liittyviä teknologioita ovat: asiakaspäässä tapahtuva tiedon suodatus, palvelinpäässä tapahtuva tiedon reititys ja asiakasprofiilien tilastollinen seuranta.
Ominaisuus 6: Automaattinen linkitys	Järjestelmä voi tiedon esitysvaiheessa reaaliajassa automaattisesti luoda linkkejä muuhun sisällöllisesti samankaltaiseen aineistoon.

Ominaisuus 7: Ryvästäminen	Ryvästämisellä tarkoitetaan aineiston jakamista toisistaan poikkeaviin luokkiin, joita ei kuitenkaan ole ennalta määrätty millään sanastolla. Ryvästys voidaan toteuttaa joko täysin automaattisesti, jolloin ohjelmisto päättelee aineistosta kuinka moneen ryppääseen ja kuinka monitasoiseen hierarkiaan kyseinen aineisto kannattaa jakaa tai sitten voidaan ennalta määrittää tehtävien rypäiden ja hierarkiatasojen lukumäärä. Aineiston ryvästämisellä voidaan tiedon esitysvaiheessa tarjota esimerkiksi päivän uutistarjonnasta lukijalle nopeasti visuaalisesti kokonaisvaltainen näkemys.
Ominaisuus 8: Automaattinen tiivistäminen	Järjestelmä palauttaa lyhyet tiivistelmät kaikista haun perusteella löytyneistä dokumenteista. Tiivistelmissä näkyy muutama lause, jotka kuvaavat haun tuloksena olevaa dokumenttia. Lauseet voivat olla tekstin eri osista.

Luokittelutesti

Testissä oli päämääränä asettaa automaattiset luokittelujärjestelmät paremmuusjärjestykseen vertailemalla automaattisten luokittelijoiden ehdotuksia toimittajien käsin tekemiin luokituksiin. Järjestelmäntoimittajat tuottivat omalla luokittelijallaan testimateriaalin uutisille omasta mielestään optimoidun listan järkeviä luokkia ja listan kymmenestä relevanteimmasta luokasta.

Saannin ja tarkkuuden arvot laskettiin erikseen järjestelmäntarjoajien antamista optimoiduista tuloslistoista, sekä listoista, jotka oli kunkin uutisen kohdalla katkaistu aina samanmittaisiksi, kuin lista STT:n alkuperäisistä luokituksista.

Tuloksista paremmuusjärjestys selviää vertailemalla eri järjestelmien saamia arvoja saannin ja tarkkuuden yhtäsuuruuspisteessä (Taulukko 8). Koska saannin ja tarkkuuden arvot eivät ole täysin samat tässäkään pisteessä käytettään jatkossa yhtäsuuruuspisteessä olevia saannin arvoja.

Optimoiduista tuloslistoista lasketuista arvoista nähdään, että yhtä lukuun ottamatta kaikki luokittelujärjestelmät saivat paremman arvon saannista, kuin tarkkuudesta, mikä tarkoittaa käytännössä sitä, että ne ehdottivat mieluummin liikaa, kuin liian vähän luokkia.

Suuret kansainväliset yritykset saivat tässä testissä selvästi kotimaisia järjestelmäntarjoajia huonommat tulokset. Käsin tehtyyn luokitteluun verrattaessa parhaiten pärjäsi Ohjelmisto 3, jonka saanti oli 47%, tosin eroa toiseksi tulleeseen Ohjelmistoon 2 (42%) oli vain alle viisi prosenttiyksikköä, mikä on testausjärjestelyiden puitteessa täysin merkityksetön ero. Kolmanneksi pääsi Ohjelmisto 4, joka sai 32% ja huonoimmin pärjäsi Ohjelmisto 1, jonka saanti oli vain 17%.

Taulukko 8. Saannin ja tarkkuuden arvot vertailtaessa luokittelujärjestelmien ehdottamia luokituksia toimittajien käsin tekemiin luokituksiin.

	Saannin ja tarkkuuden yhtäsuuruuspisteessä		**Optimoiduista tuloslistoista laskettuna**	
	Saanti %	Tarkkuus %	Saanti %	Tarkkuus %
Ohjelmisto 1	16,72	15,85	25,44	11,84
Ohjelmisto 2	41,73	40,16	56,94	20,02
Ohjelmisto 3	46,95	45,09	59,41	19,19
Ohjelmisto 4	31,62	29,99	25,71	36,26

Luokitteluehdotusten relevanssin tarkistus

Relevanssin tarkistuksessa oli päämääränä selvittää, mikä luokittelujärjestelmä on osannut ehdottaa eniten sellaisia relevantteja luokkia, joita toimittaja ei ole luokittelua tehdessään huomioinut. Tuloksista tämä selviää vertailemalla *"kaikilla relevanteilla luokilla"* laskettuja lukuarvoja *"käsin luokitelluilla luokilla"* laskettuihin. Mikäli ensin mainittu lukuarvo on suurempi tarkoittaa se, että kyseisen järjestelmän ehdottamista luokista on valittu tarkistusluennassa useampia mukaan laajennetulle ns. oikeiden luokkien listalle.

Relevanssin tarkistuksessa kaikki lukuarvot on laskettu saannin ja tarkkuuden yhtäsuuruuspisteessä. Taulukossa 9 näkyvien käsin luokitelluilla luokilla laskettujen lukuarvojen pitäisi periaatteessa olla vertailukelpoisia edellisessä kappaleessa saannin ja tarkkuuden yhtäsuuruuspisteessä laskettujen tulosten kanssa. Koska saadut lukuarvot ovat kuitenkin kauttaaltaan suurempia, voidaan päätellä, että relevanssin tarkistukseen mukaan otetussa kolmenkymmenen uutisen otoksessa korostuu helposti luokittuvien uutisten osuus. Merkittävää on kuitenkin, että eri luokittelujärjestelmien paremmuusjärjestys säilyy entisellään.

Kun uutisten lukemisen yhteydessä valittiin, mitkä kaikki ehdotetuista luokista on relevantteja, käsiteltiin STT:n luokkia ja kaikkien luokittelujärjestelmien ehdotuksia yhtäläisesti. Joissain tapauksissa kävi myös niin, että alkuperäinen STT:n luokka tulkittiin epärelevantiksi. Tämä kaikkien järjestelmien yhteiskäsittely johtaa siihen, että saadut lukuarvot ovat merkitseviä vain suhteessa toisiinsa, eikä absoluuttisina lukuarvoina.

Taulukko 9. Tulokset luokitteluehdotusten relevanssin tarkistuksesta. Käsin luokitelluilla luokilla lasketuissa tuloksissa on verrattu toimittajien tekemiä luokituksia automaattisten luokittelijoiden antamiin ehdotuksiin. Kaikilla relevanteilla luokilla lasketuissa tuloksissa on verrattu automaattisten luokittelijoiden tekemiä ehdotuksia listaan, johon on toimittajan tekemien luokitusten lisäksi otettu mukaan kaikki sellaiset järkevät luokitukset, joita jokin vertailussa mukana ollut automaattinen luokittelija ehdotti.

	Käsin luokitelluilla luokilla		**Kaikilla relevanteilla luokilla**	
	Saanti %	Tarkkuus %	Saanti %	Tarkkuus %
Ohjelmisto 1	25,81	25,81	26,24	26,36
Ohjelmisto 2	49,46	49,46	46,15	48,11
Ohjelmisto 3	52,69	53,26	61,99	63,13
Ohjelmisto 4	34,41	34,41		

Luokitteluehdotusten relevanttiuden tarkastuksessa voitiin todeta, että Ohjelmisto 3 oli ainoa, jonka käsin luokituksesta poikkeavissa ehdotuksissa oli joukossa merkittävä määrä muita relevantteja ehdotuksia. Pienellä otoksella Ohjelmisto 3:n saanti parani noin yhdeksän prosenttiyksikköä, kun ns. oikeina vertailuluokituksina käytettiin tarkastettuja relevantteja ehdotuksia toimittajien valitsemien ehdotusten sijasta. Tässä testissä Ohjelmisto 4 ei ollut mukana myöhässä toimitetun materiaalin takia.

Kahden muun järjestelmän tuloksissa ei tapahtunut merkittävää muutosta tarkistettuja luokkia käytettäessä. Se, että niiden tulokset eivät muuttuneet tarkoittaa käytännössä sitä, että muutamia niiden ehdottamiaan luokkia tuli valituksi relevanttien luokkien listalle, mikä paransi niiden tarkkuutta, mutta vastaavasti muiden järjestelmien ehdottamia luokkia oli saman verran, joka puolestaan heikensi niiden saantia.

7. Johtopäätökset ja kehitysehdotukset

Tutkimuksen alussa esitettiin ajatus, että mukaan STT:n journalistisesti tuottamasta uutismateriaalista on mahdollista haku- ja personointiteknologioita hyödyntämällä jalostaa uusia entistä paremmin käyttäjien tarpeita vastaavia uutispalveluita. Tutkimuksen tavoitteena oli selvittää, miten haku- ja personointiteknologioita voidaan hyödyntää uutispalveluiden tuottamisessa. Johtoajatuksesta muokattiin kolme tutkimuskysymystä:

1.) Miten uutistoimistojen ja muiden uutislähteiden nykyisiä palveluita käytetään kohderyhmän arkityössä ja minkälaisia parannustarpeita niihin liittyy?

2.) Minkälaisia haku- ja personointiteknologioita on tarjolla ja minkälaisia ominaisuuksia kaupallisilla ohjelmistoilla on?

3.) Miten haku- ja personointiteknologioilla voidaan tuottaa uutispalveluja helpottamaan loppukäyttäjien työtä asiakasorganisaatioissa?

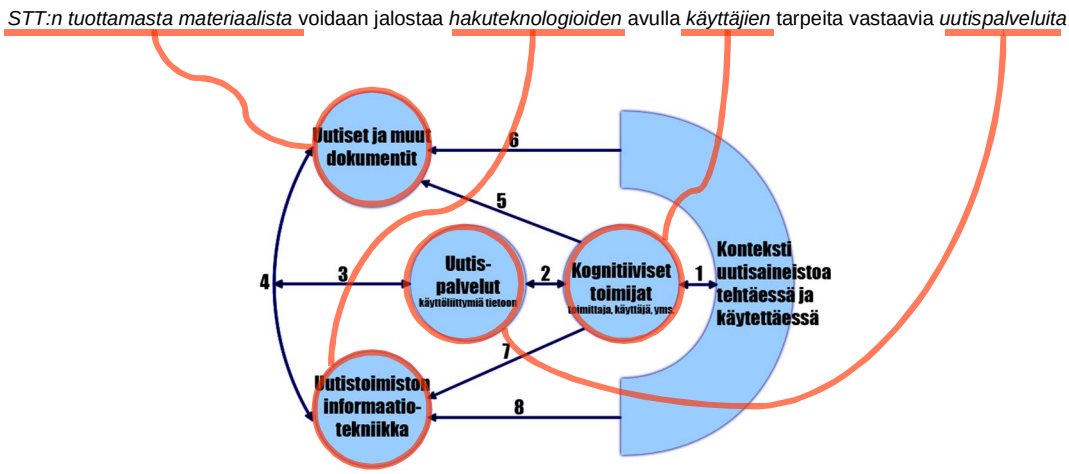
Kahdella ensimmäisellä kysymyksellä saadaan välttämättömiä ennekkotietoja, jotta pystyttäisiin vastaamaan viimeiseen kysymykseen, jossa kiteytyy tämän tutkimuksen tarkoitus.

Ensimmäiseen kahteen kysymykseen vastattiin kohderyhmään kuuluvien loppukäyttäjien haastattelututkimuksella, sekä kaupallisten haku- ja personointiteknologian ohjelmistojen vertailulla. Tässä luvussa esitetään ensin STT:n informaatiojärjestelmän malli (kappale 7.1.), joka jäsentää haastattelututkimuksen ja ohjelmistovertailun yhdeksi kokonaisuudeksi, sen jälkeen tehdään yhteenveto haastatteluiden ja teknologiakartoituksen tuloksista kappaleessa 7.2 sekä lopuksi kappaleessa 7.3 annetaan kolme kehitysehdotusta, jotka yhdessä vastaavat viimeiseen tutkimuskysymykseen.

7.1. STT:n informaatiojärjestelmä

Ingwersenin ja Järvelinin mallin (kappale 2.4.3.) sovellus kohdeyritykseen tuotti mallin, joka kuvaa STT:n toimintaa informaatiojärjestelmänä. Informaatiojärjestelmän malli toimi yleisenä jäsennyksenä koko tutkimukselle.

Valittu tiedonhankinnan ja tiedonhaun yhdistetty malli tuki hyvin STT:n informaatiojärjestelmän kokonaisvaltaista tarkastelua. Koska loppukäyttäjien tarpeet eivät olleet entuudestaan tunnettuja, eikä myöskään järjestelmän teknisiä ominaisuuksia haluttu kiinnittää ennakkoon olisi pelkästään järjestelmäkeskeisen tai pelkästään käyttäjäkeskeisen mallin soveltaminen antanut liian suppean kuvan.



Kuva 18. Tutkimuksen lähtökohta jäsennettynä STT:n informaatiojärjestelmää kuvaavan mallin mukaisesti..

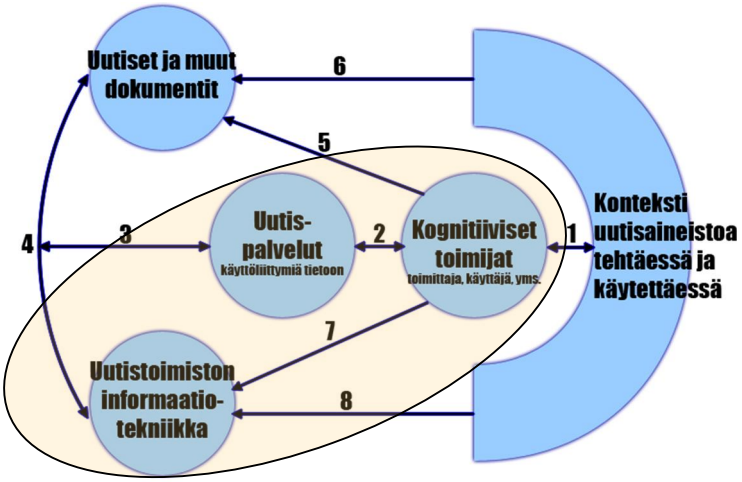
Työn lähtökohtana ollut ajatus, että *STT:n tuottamasta materiaalista* voidaan jalostaa *hakuteknologioiden* avulla *käyttäjien* tarpeita vastaavia *uutispalveluita*, jäsentyy informaatiojärjestelmän mallin mukaisesti. Tällöin STT:n tuottama materiaali vastaa informaatiojärjestelmän mallissa komponenttia *Uutiset ja muut dokumentit*, hakuteknologiat vastaavat komponenttia *Uutistoimiston informaatiotekniikka*, käyttäjät vastaavat komponenttia *Kognitiiviset toimijat* ja *Uutispalvelut* esiintyvät mallissakin samalla nimellä. Mallin viides komponentti, *Konteksti uutisaineistoa tehtäessä ja käytettäessä*, ei suoraan liity työn taustalla olevaan johtoajatukseen, mutta sekin tuli huomioitua haastattelututkimuksen yhteydessä.

Ensimmäinen tutkimuskysymys: ”Miten uutistoimistojen ja muiden uutislähteiden nykyisiä palveluita käytetään kohderyhmän arkityössä ja minkälaisia parannustarpeita niihin liittyy?” keskittyy mallissa keskellä ja vasemmalla oleviin komponentteihin *Uutispalvelut*, *Kognitiiviset toimijat* ja *Konteksti*. Tähän tutkimuskysymykseen etsittiin vastauksia tekemällä haastattelututkimus (katso Kuva 16), jossa haastateltavat olivat palveluiden loppukäyttäjiä ja haastatteluteemat luotasivat käytettyjä palveluita, niiden käyttötapoja sekä kontekstuaalisia tekijöitä.

Toinen tutkimuskysymys: ”Minkälaisia haku- ja personointiteknologioita on tarjolla ja minkälaisia ominaisuuksia kaupallisilla ohjelmistoilla on?” keskittyi erityisesti mallissa *Informaatiotekniikka* komponenttiin. Tähän tutkimuskysymykseen vastattiin tekemällä ohjelmistovertailu viidelle kaupalliselle hakuohjelmistolle. Ohjelmistovertailussa hakujärjestelmien ominaisuuksia kartoitettiin ja verrattiin STT:n uutispalveluiden tuotannosta nousseisiin tarpeisiin. Ohjelmistojen luokitteluominaisuutta testattiin myös STT:n omalla materiaalilla. Näin ollen ohjelmistovertailussa huomioitiin informaatiojärjestelmän mallissa vasemmalla ja keskellä olevat komponentit Uutiset ja muut dokumentit, Uutispalvelut ja Uutistoimiston informaatiotekniikka (katso Kuva 17).

Yhdessä haastattelututkimus ja ohjelmistovertailu kattavat kaikki komponentit STT:n informaatiojärjestelmän mallissa. Molemmissa osatutkimuksissa kerättiin tietoa Uutispalveluista. Uutispalvelut onkin keskeinen komponentti kolmannessa ja koko tutkimuksen kannalta tärkeimmässä tutkimuskysymyksessä: ”Miten haku- ja personointiteknologioilla voidaan tuottaa uutispalveluja helpottamaan loppukäyttäjien työtä asiakasorganisaatioissa?”. Kehitysehdotukset ovat uusia

palveluita ja parannuksia entisiin palveluihin, eli informaatiojärjestelmän mallissa ne kohdistuvat ”Uutispalvelut” komponenttiin. Kehitysehdotukset vastaavat loppukäyttäjien (kognitiivisten toimijoiden) tarpeisiin ja perustuvat informaatiotekniikan (uutistoimiston informaatiotekniikka) suomiin mahdollisuuksiin.



Kuva 19. Kehitysehdotukset kohdistuvat STT:n informaatiojärjestelmän "Uutispalvelut" komponenttiin, mutta ne perustuvat sekä kognitiivisten toimijoiden eli uutispalveluiden käyttäjien tarpeisiin että informaatiotekniikan suomiin mahdollisuuksiin.

7.2. Yhteenveto

Tässä kappaleessa vedetään yhteen kaksi osatutkimusta ja tehdään tarpeelliset johtopäätökset saaduista tuloksista. Haastattelututkimuksilla selvitettiin uutispalveluiden käyttöä kohderyhmässä, erityisesti uutispalveluiden käyttötarkoitusten jäsennys palvelee kehitysehdotusten tekemistä. Ohjelmistovertailussa selvitettiin markkinoilla olevien uutisten luokitteluun kykenevien ohjelmistojen ominaisuuksia ja testattiin ohjelmistojen luokitteluominaisuuden toimivuutta STT:n omalla materiaalilla. Kehitysehdotukset pohjautuvat niihin ominaisuuksiin, joita vertailun ohjelmistoilla oli. Varsinaisen luokittelutestin tuloksia ei hyödynnetä kehitysehdotuksia tehtäessä.

7.2.1. Uutispalveluiden käyttö

STT:n nykyisillä palveluilla on laaja pääasiassa media-asiakkaista muodostuva käyttäjäkunta. Laajin kasvupotentiaali asiakaskunnassa on kuitenkin median ulkopuolisten asiakkaiden ryhmässä, joten uusia uutispalveluita pyritään kehittämään erityisesti tälle asiakasryhmälle. Kohderyhmäksi tutkimukseen valittiin nykyiset ja potentiaaliset median ulkopuoliset STT:n asiakkaat.

Tutkimuskysymyksestä johdettiin neljä haastatteluteemaa: palvelut, käyttötarkoitukset, käyttötilanteet ja kehityskohteet. Tutkimusta varten tehtiin kahdeksan teemahaastattelua kohderyhmään kuuluville STT:n nykyisten palveluiden käyttäjälle.

Tehdyt kahdeksan haastattelua ovat tapaustutkimuksia, jotka antavat arvokasta tietoa ja viitteitä palveluiden jatkokehittelyyn. Palvelukehityksen myöhemmässä vaiheessa

paremman yleistettävyyden saavuttamiseksi on palvelukonsepteja syytä testata laajemmalla otoksella.

Kuva 20. Haastatteluiden tuloksena syntynyt uutispalveluiden käyttötarkoitusten jäsennys.

Haastatteluaineiston valossa voidaan todeta median ulkopuolisilla STT:n asiakkailla olevat neljä pääasiallista toisistaan erotettavaa käyttötarkoitusta uutistoimistojen ja muiden uutislähteiden käytölle:

Uutisseuranta on pääasiallinen käyttötarkoitus STT:n palveluille. Uutisia seurataan organisaatioissa, jotta niihin voidaan reagoida tarvittaessa tai varaudutaan siihen, että julkisuudessa esitetty tieto vaikuttaa organisaation toimintaan.

Mediaviestintä korostui STT:n palveluiden käyttötarkoituksena median ulkopuolisilla asiakkailla. STT:tä käytetään mediaviestinnän kanavana ja toisaalta seurataan muiden organisaatioiden viestintää ja omien viestien läpimenoa.

Tiedonhaku oli nykyisellään vähäisempi käyttötarkoitus, sillä uutismateriaalin käytössä korostuu ajallinen tuoreus ja osittainen sisällön ennalta arvaamattomuus. Myös jatkuvasti päivittyvästä materiaalista voidaan hakea tietoa tietyillä kriteereillä, mutta STT:n nykyiset palvelut eivät sitä tue.

Julkaisutoiminta on STT:n palveluiden pääasiallinen käyttötarkoitus media-asiakkaille. Vähäisemmässä määrin myös median ulkopuoliset asiakkaat käyttävät STT:n aineistoa lähteenä omiin julkaisuihinsa.

Edellä mainituista uutisseuranta ja mediaviestintä olivat merkittävimmässä asemassa ja niitä voitiin haastatteluaineiston pohjalta jaotella edelleen tarkemmin. Uutisseurantaan johtavia syitä olivat: tarve reagoida organisaation kannalta merkittäviin uutisiin, tarve varautua siihen, että jokin uutinen saattaa vaikuttaa organisaation toimintaan. Mediaviestintään liittyvä uutispalveluiden käyttö puolestaan jakautui: mediaviestinnän suunnitteluun ja kehittämiseen, tiedottamiseen ja viestien läpimenon sekä julkisuuskuvan seurantaan. Puhdas tiedonhaku ja julkaisutoiminta, joissa uutispalveluita hyödynnettäisiin, oli vähäisempää median ulkopuolisten asiakkaiden keskuudessa. Nämäkin käyttötarkoitukset tulivat esille ja on siten syytä huomioida uusien palveluiden kehitystyössä.

Käyttötapoihin liittyen haastatteluissa tuli ilmi, että yleensä uutisia selaillaan ja seurataan rutiininomaisesti muun työn ohella. Seuranta ei kuitenkaan usein ole kovin systemaattista, vaan jos tyypillinen rutiini esimerkiksi kiireisen päivän takia ei toteudu, lukematta jääneisiin uutisiin ei enää palata myöhemmin. Teemahaastatteluiden tuloksena hahmottui kuva siitä, miten STT:n palveluita käytetään muualla, kuin lehtitaloissa, mitä nämä median ulkopuoliset asiakkaat arvostavat nykyisissä palveluissa ja mihin toivotaan muutoksia.

Haastatteluiden lomassa tuli myös esille jonkin verran STT:n nykyisiin palveluihin liittyviä kohdennettuja parannusehdotuksia, kuten arkistohaun nopeuden parantaminen ja mahdollisuus kirjautua yhdellä salasanalla kaikkiin palveluihin.

Suurin osa käyttäjistä ei ainakaan tietoisesti osaa toivoa mitään edistyksellisempiä hakuominaisuuksia, kuten samankaltaisten uutisten hakua, erilaisia osasto-, aihe- ja aikarajauksia tai hienojakoisempaa luokittelua. Poikkeuksena ovat enemmän arkistopalvelua käyttävät käyttäjät, jotka toivoivat kattavampia mahdollisuuksia kohdentaa ja rajata hakuja. Yleisesti ottaen loppukäyttäjät eivät osanneet esittää toivomuksia kokonaan uusista palveluista, mitä ei vielä ole tarjolla. Käyttötapoja ja suoria parannusehdotuksia ei vähäisen haastatteluista kertyneen materiaalin takia tässä työssä analysoitu systemaattisesti, mutta nekin on huomioitu kehitysehdotuksia tehtäessä. Kehitysehdotukset löytyvät kappaleesta 7.3

7.2.2. Hakuohjelmistojen ominaisuudet

Haku- ja personointiteknologioilla tarkoitetaan väljästi kaikkia niitä tietoteknisiä ratkaisuja, joilla voitaisiin automaatioon perustuen kustannustehokkaasti tuottaa kohderyhmälle hyödyllisiä uutispalveluita STT:n uutismateriaalista.

Tutkimukseen lähdettäessä Suomen Tietotoimistolla oli jo kontakteja joihinkin ohjelmistotaloihin ja alustavia ideoita siitä, minkälaisista uusista palveluista voisi olla kysymys. STT:n sisäisissä palavereissa selvitettiin tarpeet ja toiveet uusia teknologioita kohtaan ja täsmennettiin ne kuudeksi tarvemäärittelyksi. Tarvemäärittelyitä tehtäessä ilmeni, että STT:n tavoitteiden kannalta merkittävin yksittäinen ominaisuus, jota haku- ja personointiteknologioilta odotetaan, on kyky luokitella automaattisesti uutisaineistoa ennalta määrättyihin luokkiin.

Tarvemäärittelyiden pohjalta valittiin ohjelmistovertailuun mukaan otettavat ohjelmistot. Ohjelmistovertailuun valittiin viisi ohjelmistoa, jotka kykenivät automaattisesti luokittelemaan uutisaineistoa. Ohjelmistontuottajien tarjontaan tutustuttiin taustakartoituksen, yritystapaamisten ja demopäivien muodossa. Yhteisenä piirteenä ohjelmistoilla oli luokitteluominaisuus, mutta sen ohella ne erosivat toisistaan merkittävästi niin ominaisuuksiltaan, tekniikoiltaan kuin hinnaltaankin. Vertailtujen ohjelmistojen ominaisuuksien kirjosta identifioitiin kahdeksan STT:n toiminnan kannalta relevanttia ominaisuutta, joidenka varaan kehitysehdotukset voidaan rakentaa. Nämä ominaisuudet valittiin sillä perusteella, että niitä oli useammassa kuin yhdessä vertailluista ohjelmistoista ja ne ovat relevantteja aiemmin tehtyjen tarvemäärittelyiden suhteen.

Taulukko 10. STT:n hakuteknologioihin kohdistuvat tarpeet ja hakuohjelmistojen ominaisuudet. Tarpeet ja ominaisuudet on esitetty laajemmin taulukoissa 4 ja 7.

Tarpeet	Tarve 1: Sisään tulevan materiaalin automaattinen esikäsittely
	Tarve 2: Toimitetun materiaalin luokittelun automatisointi
	Tarve 3: Yritysnimien merkinnän automatisointi
	Tarve 4: Muun materiaalin luokittelu
	Tarve 5: Aineiston automaattinen yhdistely
	Tarve 6: Asiakaskohtaisesti eriytettyjen uutispalveluiden tuottaminen

Ominaisuudet	Ominaisuus 1: Nimien automaattinen merkitseminen
	Ominaisuus 2: Automaattinen luokittelu
	Ominaisuus 3: Luokittelusanaston muokattavuus
	Ominaisuus 4: Hakuominaisuudet
	Ominaisuus 5:Sisällön eriyttäminen
	Ominaisuus 6: Automaattinen linkitys
	Ominaisuus 7: Ryvästäminen
	Ominaisuus 8: Automaattinen tiivistäminen

Alun perin viidestä ohjelmistovertailussa mukana olevasta ohjelmistosta valittiin niiden tarjoamien ominaisuuksien perusteella neljä potentiaalisinta ohjelmistoa jatkotestiin, jossa selvitettiin, kuinka hyvin nimenomaan luokitteluominaisuus toimii käytettäessä STT:n omaa aineistoa.

Luokittelutestissä mikään järjestelmä ei saavuttanut STT:n tarpeisiin nähden riittävää riittävän tarkkuustasoa. Saannin ja tarkkuuden olivat parhaimmillaankin alle 50% luokkaa. On kuitenkin huomioitava, että testausjärjestelyssä ei huomioitu sitä mekaniikkaa, jolla toimittajan valitsemat luokitustiedot tallentuvat toimitusjärjestelmässä. Tämän takia saatuja lukuarvoja sellaisenaan ei voida pitää riittävän luotettavina arvioimaan luokittelun todellista onnistumista, vaan niiden avulla voidaan ainoastaan verrata nyt testattuja järjestelmiä keskenään ja todeta niiden paremmuusjärjestys.

Parempiin tuloksiin päästäneen mm. parantamalla opetusaineiston laatua ja määrää. Voidaankin todeta, että STT:n käsinluokiteltu arkistomateriaali ei sellaisenaan kelpaa tuotantokäyttöön tarkoitetun järjestelmän opetusaineistoksi.

Todennäköisesti Ohjelmisto 3, joka ei käytä luokittelussa tilastollisia metodeja pärjäsi tässä testissä näinkin hyvin nimenomaan sen takia, että tilastollisuuteen perustuvat metodit kangistuivat heikkolaatuisen opetusdatan takia.

Tutkimuksen edetessä ja heikkojen luokittelutulosten ja muiden ongelmien paljastuessa olen joutunut useampaan kertaan kyseenalaistamaan sen, tarvitaanko

välttämättä johonkin tiettyyn sanastoon perustuvaa luokittelua, vai ajaisiko hyvin toimiva hakukone saman asian.

Jatkotoimenpiteenä automaattiseen luokitteluun liittyen suosittelen, että selvitetään se, mihin automaattista luokittelua täsmälleen tarvitaan. Sitä kautta voidaan määrittää ns. riittävä taso luokittelun laadulle, eli saannille ja tarkkuudelle. Kun tavoitetaso on tiedossa, voidaan valita STT:n kokonaistietoratkaisun kannalta sopivin järjestelmä lopulliseen testaukseen. Lopullisessa testauksessa selvitetään, onko automaattinen luokittelu teknologiana riittävän kypsää, jotta luokittelu voidaan jättää kokonaan koneen huoleksi.

Luokittelutestin tuloksia ei hyödynnetä kehitysehdotusten tekemisessä, sillä kehitysehdotukset ovat yleisellä tasolla eivätkä perustu minkään tietyn järjestelmän ominaisuuksiin.

7.3. Kehitysehdotukset

Käyttäjien tarpeiden ja hakuteknologioiden mahdollisuuksien pohjalta annetaan tässä kolme kehitysehdotusta, joita voidaan hyödyntää STT:n informaatiojärjestelmän suunnittelun tukena. Parannusehdotukset kohdistuvat informaatiojärjestelmän (Kuva 15) *uutispalvelut* ja *uutistoimiston informaatiotekniikka* komponentteihin.

Palveluiden parannustoiveita tuli käyttäjähaastatteluissa esille rajallisesti, joten nämä kehitysehdotukset ovat minun tulkintojani siitä, miten olemassa oleva teknologia voisi tukea kohderyhmän toimintaa. Konseptisuunnittelussa olennaista on iteratiivisuus, nyt esitettyjä parannusehdotuksia voidaan ja pitäisikin evaluoida kohderyhmän kanssa ja kehittää sen jälkeen eteenpäin.

Olen muokannut kolme parannuskokonaisuutta, jotka kattavat suuren osan haastatteluissa esille nousseista tarpeista. Teknologiakartoituksen yhteydessä olen pyrkinyt selvittämään, mikä on nykyteknologialla mahdollista. Toteutusehdotuksissa tukeudun olemassa oleviin teknologioihin.

Kunkin ehdotuksen aluksi esitetään ne käyttötarkoitukset, joita uusi palvelu tai ominaisuus ensisijaisesti tukisi, sekä muut haastatteluissa esille nousseet tarpeet, joihin se vastaa. Seuraavaksi esitetään palvelun tai ominaisuuden toiminnallinen kuvaus käyttäjän näkökulmasta. Palvelukuvaukset ovat kokonaan uusia, ellei niissä erikseen mainita, että jokin asia on toteutettu samalla tavoin , kuin nykyisissäkin palveluissa. Tämä palvelun kuvaus vastaa STT:n informaatiojärjestelmän, *Uutispalvelut: käyttöliittymiä tietoon*, komponenttia. Lopuksi esitetään tekniseltä kannalta toteutusehdotus, joka puolestaan vastaa, *uutistoimiston informaatiotekniikka*, komponenttia. Toteutusehdotuksen yhteydessä listataan ne ominaisuudet, joita palvelun toteuttaminen vaatii ohjelmistolta. Nämä ominaisuudet on poimittu ohjelmistovertailussa mukana olleiden uutisaineiston luokitteluun kykenevien ohjelmistojen ominaisuuksista.

7.3.1. Hakuominaisuudet ja selailu

Tällä hetkellä STT:n tarjoamia uutisia pääasiassa selaillaan ilman hakutoimintojen hyödyntämistä. Silloinkin kun hakuja tehdään, ne ovat suurimmaksi osaksi rutiininomaisia hakuja, joissa käyttäjällä on jo hyvin selkeästi tiedossa, mitä hän hakee ja millä hakusanoilla. Uutispalvelun seuranta on vahvasti sidoksissa käyttäjän arkirutiineihin.

Haastatelluista käyttäjistä muutama kuitenkin käytti STT:n arkistopalvelua satunnaisesti ja he ilmaisivat selkeitä toiveita mm. hakumahdollisuuksien laajentamiseen ja hakujen nopeuttamiseen.

Suurimman osan käyttäjistä käyttökokemusta voidaan parantaa hienovaraisilla uutisten selailua helpottavilla ratkaisuilla, kuten saman uutisen eri versioiden yhteen niputtamisella. mutta vaativammille käyttäjille olisi kuitenkin voitava tarjota kehittyneempiä hakuominaisuuksia.

Näkymä uutispalveluun on lista, jossa tuoreimmat uutiset näkyvät ylimpänä, kuten nykyäänkin, mutta sillä erotuksella, että samasta aiheesta kirjoitetut päivitetyt uutiset näkyvät vain kerran viimeisimmän päivityksen kohdalla (edelliset päivitykset ovat linkitettyinä tuoreimpaan). Hakutoiminto on vain tapa muuttaa tätä oletusarvoista esitysjärjestystä jonkun aiheen tai muun hakukriteerin perusteella.

Hakutuloksia esitettäessä käyttäjälle näytetään tulokset oletusarvoisesti aikajärjestyksessä, mutta annetaan sivupalkissa esimerkiksi aiheluokittelun mukaiset ryhmät ja tarjotaan mahdollisuus hakutulosten rajaamiseen ja järjestämiseen eri kriteerien perusteella.

Uutiset on myös linkitetty toisiin samankaltaisiin uutisiin uutisjutun lopussa olevalla linkkilistalla. Tämä on luonnollinen tapa tukea vakiintunutta käyttötapaa, jossa käyttäjä mieluummin selailee uutisia, kuin hakee niitä minkään tarkasti määriteltyjen hakuehtojen perusteella. Usein käyttäjä saattaa jo entuudestaan tietää tai muistaa, että STT on julkaissut jotain samaan aiheeseen liittyvää ja hän mielellään lukisi taustatietoja, mutta kynnys lähteä aktiivisesti etsimään aineistoa hakutoiminnon avulla on melko korkea.

Ohjenuorana käyttöliittymän suunnittelussa pidetään sitä, että käyttöliittymä ensisijaisesti tukee uutisten selailua ja hakutoiminnallisuudet tulevat lisänä tähän. Käyttöliittymäsuunnittelussa huomioidaan eri vaatimustason käyttäjät siten että peruskäyttöliittymässä on vain yksi hakukenttä näkyvillä ja sen vieressä on linkki, josta pääsee kehittyneempien hakutoimintojen pariin halutessaan.

Peruskäyttöliittymään lisättävä hakutoiminnallisuus edellyttää uutisaineistolla hyvin toimivaa hakukonetta. Valtaosa aineistosta on suomenkielistä, mutta hakukoneen tulee käsitellä myös ruotsin ja englanninkielisiä artikkeleita kohtuullisesti. Minimissään hakutoiminnallisuuksien toteuttaminen uutisten selailun tueksi ei vaadi kuitenkaan muuta kuin tavanomaisen sanahaun. Metadatan rikastaminen luokituksilla ja erisnimien poiminnalla tarjoaa kuitenkin käyttäjälle enemmän vaihtoehtoja, joiden perusteella rajata hakujaan. Hakutoiminnallisuuden käytettävyyttä voidaan parantaa käsitehaun avulla. Käsitehaku erottaa samalla tavalla kirjoitettuja, mutta eriasiaa tarkoittavia hakusanoja toisistaan. Ottaen huomioon, että hakutoiminnallisuus on kuitenkin pääasiassa vain tukemassa selailemalla tapahtuvaa uutisten lueskelua, ei nämä kehittyneemmät toiminnot ole välttämättömiä.

Uutisten linkittäminen tehdään automaattisesti etsimällä tietokannasta mahdollisimman samankaltaisia uutisjuttuja painottaen mm. juttujen ajallista tuoreutta. Näistä samankaltaisista uutisista koostetaan noin 5-10 linkin lista uutisjutun loppuun. Jutun julkaisusta vastuussa oleva toimittaja voi joko hyväksyä tai hylätä automaattiset linkkiehdotukset tai halutessaan lisätä manuaalisesti linkin johonkin sellaiseen uutisjuttuun, jota automaattinen linkitys ei löytänyt. Linkitys mahdollistaa myös erilaisten aineistojen, kuten päivälistamerkintöjen, kuvien, tiedotteiden ja uutisten linkittämisen toisiinsa.

7.3.2. Yksi haku koko aineistoon

Käyttäjän kannalta on epäolennaista, miten STT:n tarjoamat palvelut on rajattu, nimetty ja eroteltu toisistaan. Yleensä haastatellut aika hyvin tiesivät, mitä mikäkin palvelu sisältää, vaikka sekaannuksia joidenkin palveluiden välillä tapahtuikin.

Tiedonhaku oli STT:n nykyisten palveluiden käyttötarkoituksena vähäinen verrattuna uutisseurantaan ja mediaviestintään. Osasyynä tiedonhaun vähäiseen painoarvoon on, ettei STT:n nykyisin tarjoamissa palveluissa ole kovin laajoja hakutoiminnallisuudet. Merkittävämpi tekijä on kuitenkin, että uutiset tiedonlähteenä vastaavat yleensä orientoivaan tiedontarpeeseen, jolloin käyttäjä ei ennakkoon välttämättä tiedä, mistä aiheesta hän oikeastaan etsii tietoa.

Tarjoamalla hakupalveluita, jotka tukevat orientoivan tiedon hankkimista tuetaan suurta määrää käyttäjiä, vaikkei tiedonhaku erityisenä käyttötarkoituksena korostukaan. Tällaiset hakupalvelut ovat hyödyllisiä niin uutisseurannassa, kuin mediaviestinnässäkin.

Hakutoimintojen kannalta STT:n palveluiden yhdistäminen tarkoittaa, että yhdellä haulla voi hakea koko aineistosta. Mikäli käyttäjällä on oikeudet sekä arkistomateriaalin lukemiseen, että tuoreisiin uutisiin on hämmentävää, että hän ei voi hakea koko aineistosta kerrallaan.

Käytettävyyden parantamiseksi STT:n järjestelmään pääsee yhdellä salasanalla. Tunnistettuaan käyttäjän, järjestelmä tarjoaa pääsyn niihin ominaisuuksiin, joista asiakas on maksanut. Hakutuloksia esitettäessä tarkistetaan, mihin osaan aineistoa asiakkaalla on lukuoikeudet. Asiakkaalle luonnollisesti näytettään kaikki osumat koko siitä osasta aineistoa, johon hänellä on oikeudet, mutta myös oikeuksien ulkopuolelle jääneiden osumien olemassaolo voidaan ilmaistaan käyttäjälle esimerkiksi lukittuina kohteina, joista näkyy vain otsikko. Lukituista kohteista on tarjolla linkki ja ohjeet, miten kyseisen aineiston saa tilattua. Asiakkaille voidaan tarjota mahdollisuutta vain kyseisen artikkelin kertatilaukseen tai esimerkiksi määräaikaista lukuoikeutta artikkelikategoriaan, johon lukittu kohde kuuluu.

Ensisijaisesti hakutoiminto kohdistuu koko aineistoon, mutta käyttäjä voi halutessaan rajata hakua koskemaan esimerkiksi vain materiaalia, johon hän on ostanut lukuoikeudet, vain kuvia, vain tuoreita uutisia tms.

Kaikkea STT:n kautta tarjottua materiaalia, kuten arkistomateriaalia, uutispalveluna myytäviä tuoreita uutisia, lista- ja kalenterimerkintöjä, kuvia yms. käsitellään hakutoimintojen kannalta yhtenä kokonaisuutena.

7.3.3. Profilointi tai personointi

Profiloinnilla tarkoitetaan sitä, että uutismassasta poimitaan jollekin käyttäjäryhmälle heidän kannaltaan keskeisimmät uutiset ja toimitetaan vain ne. Profiilit voisivat olla STT:n tapauksessa asiakaskohtaisia. Puhuttaessa personoinnista tarkoitetaan, että kullakin käyttäjällä olisi henkilökohtainen profiilinsa. Usein asiakasorganisaatioissa on useita eri käyttäjiä, joilla voi olla hieman toisistaan poikkeavia tarpeita ja ideaalitapauksessa STT tarjoaisi käyttäjäkohtaista profilointia ilman lisäkustannuksia. Haastatteluiden perusteella näyttää siltä, että asiakkaat odottavat ainakin asiakaskohtaisen profiloinnin toteuttamista, mutta asiakasprofiloinnin lisäksi myös käyttäjäkohtaiselle personoinnille on kysyntää.

Uutispalveluun sisällytetään vakio-ominaisuutena ainakin yksi asiakaskohtainen profiili, joka määrittelee sen, mitkä ovat juuri tätä asiakasta kiinnostavia uutisia.

Oletusarvoinen käyttöliittymä profiloidulle uutismateriaalille on online palvelussa oleva "omat uutiset" välilehti. Halutessaan asiakas voi myös määritellä vaihtoehtoisia käyttöliittymiä, kuten uutisten toimittamisen sähköpostiin tai niiden julkaisemisen yrityksen sisäverkossa tms. Toinen vaihtoehtoinen käyttöliittymä profiilin mukaisten uutisten esittämiseen voisi olla hienovarainen korostus esimerkiksi värin tms. avulla normaalin uutisvirran seasta.

Jokaisella uutispalveluasiakkaalla on mahdollisuus muokata yksi hakulauseke, jonka täyttävät dokumentit tuotetaan automaattisesti hänen valitsemaansa käyttöliittymään. Yksinkertaisimmillaan profiilin hakuehto voisi olla vaikkapa vain muutama hakusana, kuten yrityksen nimi tai tuotemerkki, mutta mikäli uutisaineistoa on indeksointivaiheessa rikastettu luokituksilla, yritysnimien poiminnalla yms. voidaan hakuehdoista ja sitä myöten profiileista tehdä hyvinkin yksityiskohtaisia.

On luontevaa ajatella, että jokaiseen erikseen laskutettavaan palveluun kuuluu myös oma profiilinsa, jolloin esimerkiksi tekstiviestiuutisten tilaaja voi profiloida kännykkäänsä tulemaan vain kaikkein tärkeimmät uutiset ja normaaliin online käyttöliittymään sitten hieman laajemman skaalan mukaan.

Oman profiilin mukaisista uutisista muodostuu arkisto niillekin, jotka eivät osta arkistopalvelua. Käytännössä tämä voidaan toteuttaa se. asiakas saa pysyvän lukuoikeuden kaikkiin tilaamiinsa uutisiin, jotka ovat ilmestyneet sen jälkeen, kun hän on aloittanut uutispalvelun tilaamisen.

Toteutuksessa profiili perustuu tallennettuihin hakulausekkeisiin eli alertteihin. Teknistä toteutusta suunniteltaessa voidaan ajatella, että itse asiassa kaikki tarjottavat palvelut ovat tietynlaisia profiileja koko aineistoon. Tällöin käyttäjän oma profiili ja palveluprofiili toimivat päällekkäisinä suodattimina se. jos vaikka yritys nimeltä Toimiala Oy on ostanut oikeudet kotimaan uutisiin ja talousuutisiin ja määritellyt profiilissaan kaikki uutiset, joissa mainitaan yrityksen nimi. Tällöin he näkevät "Omat uutiset" osiossa kaikki uutiset, joissa mainitaan Toimiala Oy ja jotka on julkaistu kotimaan tai talousosastoilla. Mikäli Toimiala Oy:stä on uutisoitu vaikkapa urheiluosastolla voitaisiin uutisen otsikko näyttää kuitenkin "Omissa uutisissa", mutta tällöin heillä ei olisi lukuoikeutta uutiseen.

8. Yhteenveto

Tutkimuskysymysten asettelussa nojauduttiin informaatiotutkimuksen yhdistettyyn tiedonhankinnan ja tiedonhaun malliin (kappale 2.4.3.). Valitun mallin pohjalta Suomen Tietotoimisto kuvattiin informaatiojärjestelmänä (kappale 6.1.), jonka eri komponentteja ovat niin tutkimuksen kohderyhmään kuuluvat loppukäyttäjät, kuin tekniset järjestelmätkin, joihin kehitysehdotukset kohdistuivat. Mallin soveltaminen edesauttoi merkittävästi hahmottamaan STT:n informaatiojärjestelmää kokonaisuutena.

Kohderyhmän tarpeita ja uutispalveluiden käyttöä tutkittiin haastattelemalla ja hakuteknologioiden suomiin mahdollisuuksiin perehdyttiin tekemällä ohjelmistovertailu. Nämä kaksi osatutkimusta antoivat tarpeelliset lähtötiedot kehitysehdotusten tekemiseen.

Tehdyt kehitysehdotukset kohdistuvat STT:n informaatiojärjestelmän *uutispalvelut* ja *uutistoimiston informaatiotekniikka* komponentteihin. Ne on tehty tukemaan käyttäjähaastatteluissa esille nousseita uutispalveluiden pääasiallisia käyttötarkoituksia ja toteutusehdotukset pohjautuvat ohjelmistovertailussa mukana olleiden ohjelmistojen ominaisuuksiin.

Kolme kehitysehdotusta: *"yksi haku koko aineistoon, hakuominaisuudet ja selailu, profilointi tai personointi"* vastaa yhdessä koko tutkimuksen tavoitteeseen. Kehitysehdotukset ovat esimerkkejä siitä, miten hakuteknologioita voidaan hyödyntää uutispalveluiden tuotannossa.

Kehitysehdotusten voidaan ajatella olevan teknisen palvelukonseptoinnin ensimmäisen vaiheen ehdotuksia, joissa on haettu teknisesti mahdollisia ratkaisuja käyttäjäkunnan tarpeisiin. Jatkotyönä näitä ideoita pitää evaluoida laajemmalla otoksella kohderyhmästä ja niiden toimivuutta on syytä arvioida myös taloudellisesta näkökulmasta ennen tarkempaa jatkokehittelyä ja toteutusvaihetta.

Lähteet

Alaterä, A., Halttunen, K. & Sormunen, E. , *Tiedon organisoinnin ja kuvailumenetelmien perusteet*. Available: http://oppimateriaalit.internetix.fi/fi/avoimet/0viestinta/informaatiotutkimus/tiedon_organisoinnin/ [2007, 7/6/2007] .

Alkula, R. 2000, *Merkkijonoista suomen kielen sanoiksi*, University of Tampere.

Bannon, L.J. & Bødker, S. 1997, "Constructing Common Information Spaces.", *ECSCW*, pp. 81.

Belew, R.K. 2000, *Finding out about : a cognitive perspective on search engine technology and the WWW*, Cambridge University Press, Cambridge, U.K.

Belkin, N.J. 1984, "Cognitive models and information transfer", *Social Science Information Studies*, vol. 4, no. 2-3, pp. 111-129.

Belkin, N.J. & Croft, W.B. 1992, "Information filtering and information retrieval: two sides of the same coin?", *Communications of the ACM*, vol. 35, no. 12, pp. 29-38.

Belkin, N.J. & Croft, W.B. 1987, "Retrieval techniques", , pp. 109-145.

Boyd-Barrett, O. & Rantanen, T. 2002, "News Agencies in the Age of Internet" in *The Media: An Introduction*, eds. A. Briggs & B. Cobley, 2nd edn, Longman, .

Chen, C. & Hernon, P. 1982, *Information seeking : assessing and anticipating user needs*, Neal-Schuman, New York.

Cleverdon, C.W. 1967, "The Cranfield Tests on Index Language Devices", *Aslib Proceedings*, vol. 19, no. 6, pp. 173-194.

Deniston, M. 2003, , *Concept Searching Whitepaper* [Homepage of Fios Confidential], [Online]. Available: http://www.fiosinc.com/pdfFiles/concept_searching.pdf [2007, 04/23] .

Dervin, B. 1993, "Verbing communication: mandate for disciplinary invention. Journal of Communication", vol. 43, no. 3, pp. 45-54.

Dervin, B. & Nilan, M. 1986, "Information needs and uses", *Annual Review of Information Science and Technology*, vol. 21, pp. 3-33.

Ellis, D. 1996, "The dilemma of measurement in information retrieval research", *Journal of the American Society for Information Science American Society for Information Science*, vol. 47, no. 1, pp. 23-36.

Foltz, P.W. & Dumais, S.T. 1992, "Personalized information delivery: an analysis of information filtering methods", *Communications of the ACM*, vol. 35, no. 12, pp. 51-60.

Haasio, A. & Savolainen, R. 2004, *Tiedonhankintatutkimuksen perusteet*, BTJ Kirjastopalvelu, Helsinki.

Heaps, H.S. 1978, *Information Retrieval: Computational and Theoretical Aspects*, Academic Press, Inc, Orlando, FL, USA.

Hertzum, M. 1999, "Six roles of documents in professionals' work.", *ECSCW'99: Proceedings of the sixth conference on European Conference on Computer Supported Cooperative Work*Kluwer Academic Publishers, Norwell, MA, USA, pp. 41.

Hirsjärvi, S. & Hurme, H. 1988, *Teemahaastattelu*, 4. p edn, Yliopistopaino, Helsinki.

Hyvönen, E. 2005, *Miksi asiasanastot eivät riitä vaan tarvitaan ontologioita?*.

Ingwersen, P. & Järvelin, K. 2005, *The turn : integration of information seeking and retrieval in context*, Springer, Dordrecht.

Järvelin, K. 1995, *Tekstitiedonhaku tietokannoista : johdatus periaatteisiin ja menetelmiin*, Suomen atk-kustannus, Espoo.

Järvelin, K. & Sormunen, E. 1999, "Dokumentit kateissa? Tiedon tallennus ja haku avuksi" in , ed. I. Mäkinen, Tiedon tie : johdatus informaatiotutkimukseen edn, BTJ Kirjastopalvelu, Helsinki, pp. 110-143.

Kansalliskirjasto , *YSA (Yleinen suomalainen asiasanasto)*. Available: http://www.kansalliskirjasto.fi/kirjastoala/asiasanastot/ysa.html [2007, 8/2/2007] .

Metsämuuronen, J. 2003, *Tutkimuksen tekemisen perusteet ihmistieteissä*, International Methelp, Helsinki.

Robins, D. 2000, *Interactive Information Retrieval: Context and Basic Notions*.

Salton, G. 1989, *Automatic text processing: the transformation, analysis, and retrieval of information by computer*, Addison-Wesley Longman Publishing Co., Inc, Boston, MA, USA.

Savolainen, R. 2000, "Tiedontarpeet ja tiedonhankinta" in *Tiedon tie : johdatus informaatiotutkimukseen*, 4th edn, BTJ Kirjastopalvelu, Helsinki, pp. 73-109.

Schütze, H., Hull, D.A. & Pedersen, J.O. 1995, "A comparison of classifiers and document representations for the routing problem", *SIGIR '95: Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*ACM Press, New York, NY, USA, pp. 229.

STT 2006, , *STT - Suomen Tietotoimisto*. Available: http://www.stt.fi/fi/ [2008, 5/23] .

Syrjälä, L., Syrjäläinen, E., Ahonen, S. & Saari, S. 1994, *Laadullisen tutkimuksen työtapoja*, Kirjayhtymä, Helsinki.

Taylor, R.S. 1968, "Question-negotiation and information seeking in libraries", *College and Research Libraries*, vol. 29, pp. 178-194.

Vakkari, P. 1999, "Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval", *Information Processing & Management*, vol. 35, no. 6, pp. 819-837.

Wilson, T.D. 1999, "Models in information behaviour research", *Journal of Documentation*, vol. 55, no. 3.

Wilson, T.D. 1997, "Information behaviour: an interdisciplinary perspective", *Inf.Process.Manage.*, vol. 33, no. 4, pp. 551-572.

Liitteet

Liite A

Haastatteluaineiston kommenttien luokitteluun käytetyt luokat haastatteluteemojen mukaan järjestettynä.

Teema	Luokka	Luokkaan kuuluvien kommenttien lukumäärä
Teema 1: Palvelut (Mitä?)	Uutispalvelu	77
	Arkisto	35
	Mobiiliuutiset	34
	Tiedotepalvelu	28
	Listat ja kalenterit	19
	Muiden tuottamat palvelut	14
	Toimialaseuranta	7
	Grafiikka	5
	Artikkelipalvelu	2
	Yhteensä:	**221**
Teema 2: Käyttötarkoitukset (Miksi?)	Viestintä	62
	Uutisseuranta	48
	Tiedonhaku	44
	Reagointi	26
	Median huomio	23
	Mediaseuranta	20
	Kirjoittaminen	11
	Yhteiskunnallinen	10
	Varautuminen	9
	Yhteensä:	**253**
Teema 3: Käyttötilanteet (Miten?)	Rutiinit	82
	Tausta	32
	Uutislähteet	30
	Vaativuus	9
	Työtehtävien	1
	Yhteensä:	**154**
Teema 4: Kehityskohteet	Ominaisuudet	36
	Käytettävyys	25
	Profilointi	16
	Liitetiedostot	3
	Linkitys	2
	Yhteensä:	**82**

Liite B

Esimerkkitiedosto opetusmateriaalista (25162064.xml)

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

- <UUTINEN>

<ID>**25162064**</ID>

<PRIO>**4**</PRIO>

<CATEGORY>**Politiikka, Ihmisoikeudet, Sosiaalikysymykset, Kodittomuus, Sosiaalikysymykset, Köyhyys**</CATEGORY>

<DEPARTMENT>**Ulkomaat**</DEPARTMENT>

<HEADLINE>**Hylätyt lapset tulevat "Kaverikotiin" roskiksista, vinteiltä ja viemäreistä**</HEADLINE>

- <LEAD>

<P>**Ukrainassa lasketaan olevan satojatuhansia kaduille hylättyjä lapsia. Virallisestikin heitä on 150 000. Tilastoissa näkyvät kuitenkin vain ne lapset, jotka ovat joutuneet tekemisiin viranomaisten kanssa.**</P>

****

Tästä välistä on poistettu uutistekstiä, jotta esimerkki mahtuisi yhdelle sivulle.

****

</LEAD>

<MODIFIED>**20050105030014**</MODIFIED>

</UUTINEN>