

Yrityskäyttöön suunnitellut hakuratkaisut

# Haun sisällölytys IT- arkki- tehtuuriin

# Yrityskäyttöön suunnitellut hakuratkaisut voivat merkittävästi parantaa organisaation tuottavuutta ja tehokkuutta, jos ne otetaan käyttöön tehokkaasti.

## JOHDANTO: ARKKITEHTUURIN MERKITYS

Yrityskäyttöön suunnitellut hakuratkaisut ovat uusia tehokkaita työkaluja, joiden avulla on mahdollista saavuttaa yritysten tavoite hyödyntää tietoaan paremmin.

Monitasoisten hakuratkaisujen, jotka etsivät tietoa työpöydältä, osastoista, yrityksestä ja Internetistä, luvataan parantavan tuottavuutta ja organisaation tehokkuutta. Ratkaisujen avulla voidaan ratkaista organisaatioiden nykyisiä suuria ongelmia, kuten tarpeettomia samanlaisen tiedon säilytyspaikkoja, useita eri tiedonhakumenetelmiä ja yrityksen yhtenäisten semanttisten määritelmien puuttumista. Lisäksi niiden avulla voidaan vastata työntekijöiden kasvaviin odotuksiin siitä, että yritysjärjestelmät ovat yhtä helppokäyttöisiä ja käyttäjäystävällisiä kuin Internet-ohjelmistot ja -palvelut.

Keskeistä organisaationlaajuisen haun etujen saavuttamisessa on hakuratkaisun tehokas integrointi muihin organisaation teknologiajärjestelmiin. BearingPoint on tunnistanut organisaationlaajuisen haun tärkeimmät periaatteet ja on kehittänyt niiden pohjalta suositellun arkkitehtuurin haun käyttöönottoa varten.

## ORGANISAATIONLAAJUISEN HAUN PERIAATTEET

Hakuratkaisujen arkkitehtuurin kehittämistä ohjaavat useat yleiset periaatteet:

**Yksinkertaisuus.** Organisaationlaajuisesta hakua ohjaava periaate on pitää kaikki järjestelmät ja ratkaisut yksinkertaisina. Monimutkaisuus vähentää ohjelmistokehittäjien tuottavuutta, tekee tuotteista vaikeasti suunniteltavia, koottavia ja testattavia, aiheuttaa tietoturvaan liittyviä ongelmia ja saa loppukäyttäjät ja ylläpitäjät turhautumaan. Tieto- ja viestintätekniiikan kehittyminen tekee palvelumallista toteuttamiskelpoisen. Palveluiltaan parantuneet ohjelmistot ja yksinkertaisesti toteutettavat hakuratkaisut vastaavat käyttäjien vaatimusta houkuttelevasta, integroidusta ja toimivasta ohjelmistosta.

**Toimivuus.** Hakuratkaisun arkkitehtuurin tärkeimmät ominaisuudet ovat (tärkeysjärjestyksessä): käytettävyys, luotettavuus, turvallisuus, skaalautuvuus (käyttäjien, viestimäärien ja palveluiden mukaan), ja kyky itsenäiseen kasvuun hakuratkaisukehityksen tarjoamia uusia palveluja lisäämällä.

## TÄSSÄ NÄKÖKULMASSA:

JOHDANTO: ARKKITEHTUURIN TÄRKEYS	1
ORGANISAATIONLAAJUISEN HAUN PERIAATTEET	1
RATKAISUN KEHITYKSEN TOIMINTAPERIAATTEET	2
HAKURATKAISUN ARKKITEHTUURI	3
Käsitteellinen arkkitehtuuri	4
Esimerkki hakuratkaisun arkkitehtuurista	4
HAKURATKAISUN MAHDOLLISUUDET	8

**Luotettavuus.** Hakuratkaisun alustan on oltava arkkitehtuuriltaan luotettava. Siinä ei pitäisi käyttää viimeisintä teknologiaa. Yleissääntönä ratkaisun ydinarkkitehtuurin pitäisi käyttää tekniikkaa, joka:

- Mahdollistaa mahdollisimman suuren siirrettävyyden mihin ympäristöön tahansa.
- On testattu muissa käyttöskenaarioissa ja on todettu niissä toimivaksi.

**Tehtäväkohtaisuus.** Yrityshakuratkaisu on nähtävä tehtäväkohtaisena yrityksen komponenttina, joka on erittäin turvallinen, skaalautuva ja käytettävissä ympäri vuorokauden. Hakuratkaisun perusrakenne on luotava huomioimalla nämä attribuutit joka tasolla.

**Saavutettavuus.** Laajan käytön tukemiseksi hakuratkaisu on suunniteltava käyttömahdollisuuksiltaan laajaksi. Tämä ominaisuus on suunniteltava arkkitehtuuriin heti aluksi, sillä sitä ei voi lisätä siihen myöhemmin.

**Joustavuus.** Joustavat palvelut ovat onnistuneen hakuratkaisun tärkein tekijä. Tällä ei tarkoiteta käyttäjäpalveluita, vaan palveluita erillisinä toimintoyksikköinä, jotka ovat käytettävissä hakualustalla määriteltyjen käyttöliittymien kautta.

#### RATKAISUN KEHITYKSEN TOIMINTAPERIAATTEET

Edellä mainittuihin periaatteisiin perustuen hakuratkaisun arkkitehtuurin on noudatettava seuraavia sääntöjä:

**Avoimet standardit ja tekniikka sekä yleiset käyttöliittymät.** Hakuratkaisun on perustuttava avoimiin standardeihin, jotka edistävät tekniikan riippumattomuutta esimerkiksi J2EE- ja .NET-standardien sekä Web-palvelujen avulla. Tämä mahdollistaa mahdollisimman suuren riippumattomuuden, auttaa ehkäisemään tarvetta rajapintojen muuttamiseen toisen sovelluksen rajapinnan muuttuessa ja piilottaa ohjelmointirajapintojen (API) monimutkaisuuden järjestelmien ja sovellusten välillä.

**Modulaarinen arkkitehtuuri.** Hakuratkaisu perustuu n-kerroksiseen, komponenttipohjaiseen arkkitehtuuriin. Ratkaisussa kaikki tasot ovat abstrakteja, kaikki komponentit ovat toisistaan riippumattomia ja toiminnot on loogisesti jaettu. Arkkitehtuurin on oltava vähän riippuvuussuhteita sisältävä ja erittäin hajautettu.

**Yhteensopivuus.** XML- (extensible markup language) ja XSLT (extensible style sheet language transformation) -kielten käyttäminen luo yhteensopivuuden, joka tarjoaa abstraktion, jossa tiedon esitysmuoto eristetään alla olevasta sovelluslogiikasta. Näin tiedon esityslogiikka voidaan eristää. Tämä arkkitehtuuri perustuu toisistaan riippumattomiin laitteisiin, alustoihin ja teknologioihin.

**Metatiedon hallinta.** Metatiedon hallinta liittyy metatietosisällön luomiseen, tallentamiseen, käyttöön, vanhenemiseen ja ylläpitoon, ja siihen kuuluvat esimerkiksi jaetun metatiedon hallinnointi ja synkronointi. Yrityksen metatietojen hallinnan ansiosta organisaatio voi:

- Automatisoida metatiedon siirtämisen olemassa olevaan sisältöön. Näin hausta tulee tehokkaampaa, kun oikeat sisältöresurssit löytyvät niiden sijainnista riippumatta.
- Kehittää kapasiteettia liittämällä metatieto suoraan sisältöresursseihin niin pian niiden luomisen jälkeen kuin mahdollista.
- Luoda selkeitä organisatorisia johtamistapoja kehitystä varten ja taksonomioiden ja toimivaltuusluetteloiden jatkuvaa hallintaa.
- Ottaa käyttöön prosesseja taksonomioiden pitämiseksi ajantasaisina ja metatiedon yhteenliittymien valvomiseksi.
- Hankkia samasta sisällöstä erilaisia näkökulmia, kuten rooli, aika ja prosessi.
- Integroimalla sisältölähteitä — Hakuratkaisulla on oltava avoin arkkitehtuuri, jotta se tukee useiden sisältölähteiden integrointia. Avoimien standardien käyttöä vaaditaan, jotta alusta toimii hyvin ja turvataan useiden sisältölähteiden yhteensopivuus.

# Monitasoisten hakuratkaisujen avulla organisaatioiden toimintaa haittaavien ongelmien ratkaisu on mahdollista.

**Tietosuoja.** Yrityskäyttöön tarkoitetun hakuratkaisun on katettava yrityksen eri sisältölähteisiin liittyvät tietoturvaperiaatteet. Käyttöoikeuksien tarkistukseen ja tarkistustapaan sekä valvontaan ja henkilöllisyyden hallintaan liittyviä periaatteita on noudatettava, jotta vain asianmukaiset käyttäjät pääsevät käsiksi luottamuksellisiin tietoihin. Hakuratkaisun tietoturvan on vastattava yrityksen tietoturvakäytäntöjä ja sääntelyvaatimuksia. Joissakin tapauksissa edellytyksenä on asiakirjatason tietoturva, jotta jäsenitelemätön sisältö voidaan ottaa tehokkaaseen käyttöön.

**Verkkoinfrastruktuuri.** Ratkaisun käyttöönotossa pitää ottaa huomioon yrityksen verkkoinfrastruktuuri ja millaista asennusta se tukee. Organisaationlaajuisista hakua käytettäessä on otettava huomioon, missä etsityt tiedot sijaitsevat, ja vaikuttavatko ulkopuoliset henkilöt tai organisaatiot hakutoimintoon, turvallisuuteen ja kokonaistehokkuuteen.

Verkkosuunnittelun yleisiä näkökohtia ja sääntöjä ovat muun muassa seuraavat:

- Paljonko liikennettä haku aiheuttaa yrityksen verkossa ja siihen liittyvissä segmenteissä? Tähän suunnittelunäkökohtaan vaikuttavat indeksoitujen tietolähteiden määrä, tietolähteiden sijainti ja käytetyn verkkoinfrastruktuurin tyyppi.
- Missä hakusovelluksen käyttäjäkunta sijaitsee, samassa rakennuksessa, eri kampuksella vai suuralueverkossa (WAN)?
- Onko verkkoliikenteen luokittelu käytössä? Liikenteen luokittelu on tärkeää ympäristössä, jossa käytetään tallennusverkon (SAN) tietojen monistusta, kuten SDRF (Symmetrix Remote Data Facility) -ohjelmaa. Organisaation on otettava huomioon haun vaikutukset verkkoon ja määritettävä kunnolliset verkon segmentointikäytännöt ja QoS (Quality of Service) -implementaatiot.
- Minkätyyppinen käyttöoikeuksien valvonta on mahdollista verkossa? Monissa kokoonpanoissa hakuratkaisua käytetään yritystietojen tai muiden luottamuksellisten tietojen käyttämiseen ja tarjoamiseen, joten kunnollinen käyttöönotto on tärkeää. Tärkeitä seikkoja ovat kunnollinen verkon segmentointi, virtuaalilähiverkon (VLAN) suunnittelu, palomuurin käyttöönotto ja verkon lokitus.
- Minkälaista verkon vikasietoisuutta edellytetään? Hakuratkaisua käyttävät talous- ja henkilöstöosasto sekä muut hallinnolliset osastot, joten sitä pidetään usein tärkeänä sovelluksena, jolta edellytetään hyvää käytettävyyttä. Tämän mahdollistamiseksi hakuratkaisussa käytetään usein klusterointitekniikkaa ja erilaisia vikasietoisuusominaisuuksia.

## Tietojenpalautus ja yritystoiminnan jatkuvuus.

Tietojenpalauttamis- ja jatkuvuusperiaatteiden noudattaminen edellyttää laitteiston ja ohjelmiston vikasietoisuutta yksittäisten vikojen minimoimiseksi. Yleinen suunnittelutapa on luoda vikasietoinen infrastruktuuri käyttämällä ylimääräisiä komponentteja. Yksittäisten komponenttien vikasietoisuus kasvaa ja järjestelmän käytettävyyks sisäisille ja ulkoisille käyttäjille ja prosesseille kasvaa.

## HAKURATKAISUN ARKKITEHTUURI

Organisaationlaajuisen haun arkkitehtuurin tavoite on vankka, yrityksen tavoitteita vastaava arkkitehtuuri. Konsepti realisoituu modulaarisena, n-tasoisena yritysarkkitehtuurina, joka noudattaa edellä mainittuja standardeja ja periaatteita. Se tukee hakuratkaisua, joka vastaa nykyisiä organisaation tarpeita, tukee tulevaa kasvua, nopeuttaa tuotteiden markkinoille saamista ja pienentää omistuksen kokonaiskustannuksia.

### Käsitteellinen arkkitehtuuri

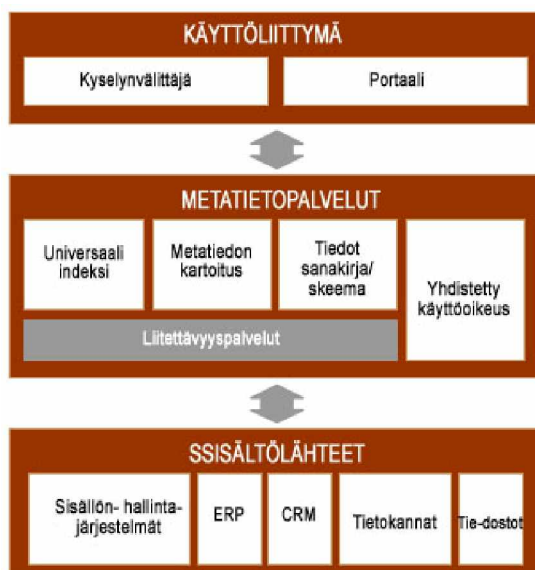
Kuvassa 1 esitetään hakuratkaisun käsitteellinen arkkitehtuuri.

Käsitteellisen arkkitehtuurin osatekijöitä ovat:

**Sisältölähteet.** Sisältölähteisiin kuuluvat yrityksen intranet, Web-sivustot, tiedostopalvelimet, sisällönhallintajärjestelmät, kuten Documentum ja FileNet sekä yrityssovellukset, jotka kattavat tapahtumaperäiset järjestelmät, kuten asiakashallinnan (CRM), toimitusketjujen hallinnan (SCM) ja yrityksen resurssisuunnittelun (ERP). Ryhmä edustaa erilaisia yrityksen sisältölähteinä toimivia sisäisiä järjestelmiä, joista käyttäjät voivat tehdä hakuja.

• **Liitettävyysspalvelut.** Liitettävyysspalvelutasoon kuuluvat sisältölähteiden perussovittimet. Taseon kuuluu kahdenlaisia sovittimia: liittimet ja yhdistetty käyttöoikeus. Liittimien avulla voidaan luoda eri tekniikkasovittimia, joita tarvitaan sisältölähteisiin liittymiseksi. Yhdistetyn käyttöoikeuden liittimillä voidaan luoda yhteys kolmannen osapuolen sisältöön.

Kuva 1. Hakuratkaisun käsitteellinen arkkitehtuuri.



**Metatietopalvelut.** Metatietopalveluiden taso tuottaa arvoa lisääviä lisäpalveluja sisällölle. Se tarjoaa kahdenlaisia palveluja: metatiedon kartoitusta ja tietojen sanakirjakartoitusta. Metatietojen kartoituspalvelun avulla voidaan määrittää eri järjestelmissä olevia yrityksen tietoja. Niiden avulla voidaan suorittaa myös yksittäistä kartoitusta yleisten elementtien välillä luomalla useista järjestelmistä yrityksen tiedoista sanakirja. Yritystietojen sanakirja esitetään metatietomääritelmänä ja tietoelementtien esiintyminä.

**Yhdistetty käyttöoikeus.** Yhdistetyn käyttöoikeuden tasolla voidaan luoda kyselyväliittäjiä, jotka hankkivat tietoja pohjalla olevilta tapahtumaperusteisilta järjestelmiltä. Yrityksen tietojen käyttäminen edellyttää universaalien hakemiston luomista asianmukaista semantiikkaa varten.

### Esimerkki hakuratkaisun arkkitehtuurista

Kuvassa 2 on esimerkki hakuratkaisun arkkitehtuurista, joka on edellä kuvatun käsitteellisen arkkitehtuurin mukainen.

Arkkitehtuurin tasot ovat:

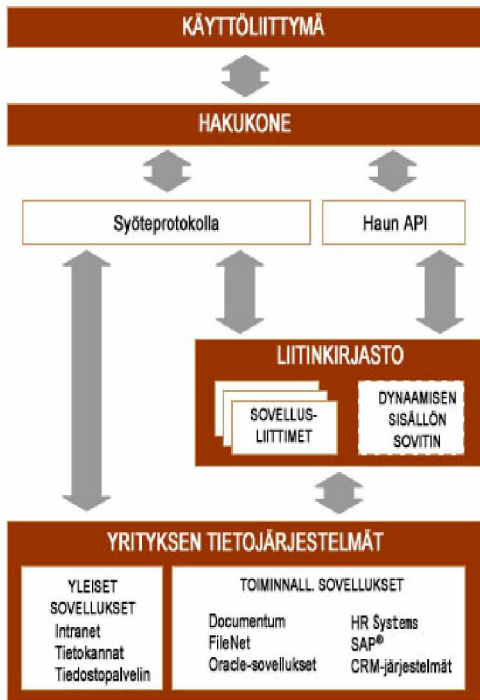
#### Käyttöliittymä

Käyttöliittymä on käyttäjävuorovaikutuksen taso, joka kontrolloi käyttäjän vuorovaikutusta Web-selaimen tai jopa kämmenlaitteen avulla. Tällä tasolla haun käyttäjille näytetään näyttöjä ja asiaanliittyviä tietoja.

Kyseessä on XML-pohjainen käyttöliittymä, jonka sovellustaso on luonut dynaamisesti käytön aikana. Hakulaitteesta riippuen XSLT-tiedosto esittää tulokset oikealla tavalla. Esittäminen ja sisältö erotetaan näin toisistaan ja vastuut jaetaan selvästi. Taso liittyy myös tiukasti turvallisuuskomponenttien käyttöoikeuksien tarkistustapaan ja käyttöoikeuksien tarkistukseen ja se varmistaa, että hakutuloksia näytetään henkilöille, joilla on riittävät valtuudet.

Tärkeitä hakuratkaisun käytettävyyttä parantavia asioita ovat roolipohjainen esitys, näytä/muokkaa- ja kontekstuaaliset ohjeltilat sekä hakutulosten sivutus.

Kuva 2. Esimerkki hakuratkaisun arkkitehtuurista



Tyypillisessä yritys ympäristössä on useita tiedosto- ja Web-sisältövarastoja intranetin uumenissa. Näissä varastoissa sisällön varsinainen isännöinti tapahtuu hyödyntämällä useita eri tekniikoita. Isännöintiin kuuluvat sisäiset turvallisuusmekanismit, yksittäiset hakemistoluettelot ja itse sisällön taustajärjestelmävarastot. Hakujärjestelmän arkkitehtuuriin on kuuluttava kolmeen eri ala-alueeseen kuuluvia toimintoja:

- **Kyselyn esikäsittelijä.** Kyselyn esikäsittelijä on keskeinen palvelu, jota voidaan kutsua useista eri paikoista. Sen tarkoituksena on hajottaa kysely osiin, tarkistaa oikeinkirjoitus, kääntää vaihtoehtoisille kielille, tunnistaa synonyymit ja tunnistaa sopivin kyselytyyppi vastaavasta sisältövarastosta, esimerkiksi avainsana-, käsite-, Boolean- tai kokotekstihaku.
- **Luettelonsovitin.** Luettelonsovitimien avulla luodaan integrointiyhteys alkuperäisen kyselyn ja taustajärjestelmän sisältöindeksiin luetteloiden kanssa. Kaikissa sisältövarastoissa yksi indeksointiominaisuus ei riitä tarjoamaan

tietosuoja-agenttien vaatimia turvallisuusominaisuuksia. Suositeltu käyttöönottopata onkin käyttää sovitimia, jotka voivat välittää kyselyn perussisältöluetteloon kyselyn tekvän henkilön puolesta.

- **Tulosten kokoaja.** Tämä taso kokoaa eri luetteloiden tulokset, jotka vastaavat hakukyselyä, ja esittää ne käyttäjälle. Tuloksissa on oltava mahdollisuus antaa palautetta palautetuista tuloksista, johtopäätöksiä vaihtoehtoisista hakulausekkeista ja mahdollisuus tehdä uusi monitasoinen tai luokiteltu haku.

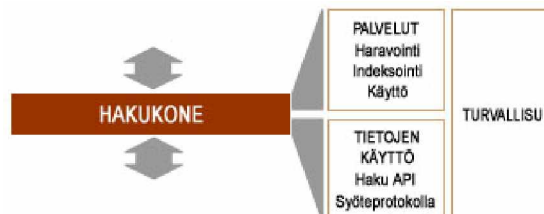
#### Hakukone

Hakukone on laitteisto- ja/tai ohjelmistotuote, jonka avulla voidaan tehdä organisaationlaajuisia hakuja yhdessä kentässä. Se hakee satoja eri tiedostomuotoja millä tahansa kielellä ja voi indeksoida miljoonia asiakirjoja. Siinä on oltava turvatoimintoja, jotta voidaan varmistaa käyttäjien näkevän vain asiakirjoja, joihin heillä valtuudet.

Hakukoneeseen on kuuluttava seuraavat tärkeimmät toiminnot (ks. kuva 3):

- **Tietojen käyttö.** Tietojen käyttö tarjoaa haun standardikäyttöliittymän, jota voidaan oletuksena isännöidä hakukoneelta. Käyttöliittymä on mukautettava sen sisältämän XSLT-tyylitaulukon muokkaamiseksi. Muut hakutoiminnot, kuten KeyMatch, vaihtoehtoiset hakusanat ja suodattimet saavat tietyt Web-sivut näkymään hakutuloksissa paremmin. Vaihtoehtoiset hakusanat ehdottavat hakukyselyille vaihtoehtoisia hakusanoja tai lausekkeita. Lisätoiminnoilla voidaan luoda loogisia tieto-osioita, joita kutsutaan kokoelmiksi ja jotka vastaavat käyttäjien erityisiin hakutarpeisiin.

Kuva 3. Hakukoneen toiminnot



- **Tietojen haravointi.** Indeksointirobotit ovat agentteja, jotka pyytävät ja hakevat tiedostoja Web-palvelimista automaattista indeksointia varten. Tietojen haravointi voi tapahtua kahdella eri tavalla: tietoja löytämällä ja tietoja palauttamalla. Määrittämyskäyttöliittymän avulla useiden tietolähteiden haravointi on mahdollista. Kaikki sisältö kerätään ja siitä luodaan päähakemisto. Hakemiston päivittäminen tapahtuu suorittamalla haravointi uudelleen.

Haravointi voidaan jakaa kolmeen vaiheeseen: Tietojen löytäminen ensimmäistä kertaa (uusia haravoitavia URL-osoitteita), indeksointi ja jatkuva haravointi. Indeksointirobotti etsii kaikista löytämistään uusista tiedostoista linkkejä. Linkit käydään läpi joko heti tai ne ajoitetaan haettaviksi myöhemmin. Indeksointirobotti on hyvä käsittelemään suojattua sisältöä ja osaa käsitellä SHTTP (secure hypertext transfer protocol) -viestintää. Hakukoneen indeksointirobotti voi neuvotella käyttöoikeuksien perustarkistuksesta, NTLM (NT LAN Manager) - tarkistuksesta sekä mukautetusta eväste- ja lomakepohjaisesta käyttöoikeudesta. Hakukoneen on haravoitava tietoa tietokannoista, kuten Oracle, SQL Server, mySQL, IBM DB2 ja Sybase. Tietotyyppi, jota indeksointirobotti ei voi käyttää, voidaan syöttää suoraan hakukoneeseen XML-muodossa.

- **Indeksointi.** Tietoja voidaan indeksoida useilla eri tavoilla etukäteen:

- Standardi- tai mukautettuja sovitteita käyttämällä hakujärjestelmään syötetty syöte.
- Käyttäminen suoraan hakujärjestelmästä (vain järjestelmän mukana toimitettu standardisovitin voi käyttää tietoja suoraan).
- HTTP- tai ei-HTTP-tietosyötteet.

- **Syötteet.** Tietyntyyppiset tiedostot kannattaa syöttää hakukoneeseen syötteiden avulla sen sijaan, että ne haetaan haravoitujen Web-sivujen linkkien kautta. Tällaisia tiedostoja ovat:

- Tiedostot, joita ei voi hakea indeksointirobotin avulla. Esimerkkejä ovat tietokannan tietueet tai ei-Web-käyttöisen järjestelmän tiedostot.
- Tiedostot, jotka voidaan indeksoida, mutta se kannattaa tehdä eri aikaan kuin hakukoneessa toimivan automaattisen indeksointiajottimen määrittäminä aikoina.

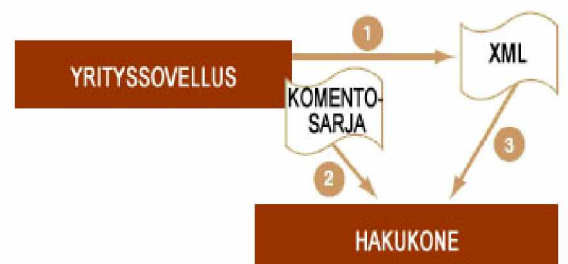
- Tiedostot, jotka voidaan indeksoida, mutta niissä ei ole Web-sivustolinkkejä, joiden avulla indeksointirobotti löytää ne uuden indeksoinnin aikana.
- Tiedostot, jotka voidaan indeksoida, mutta niiden lataaminen on nopeampaa syötteitä käyttämällä Web-palvelimen tai verkko-ongelmien takia.

Mahdollisia syötetyyppejä ovat seuraavat:

- **Sisältösyötteet**—Sisältösyötteet sisältävät sekä URL-osoitteita että niiden sisältöä. Ne voivat myös sisältää metatietoa, kuten tietueen viimeisen muokkaamisajankohdan. Sisältösyötteelle voidaan määritellä mikä tahansa tietolähteen nimi.
- **Web-syötteet**—Web-syötteessä kaikki tietueet sisältävät URL-osoitteita mutta eivät sisältöä. Näitä URL-osoitteita indeksoidaan normaaliin tapaan. Tällaisissa syötteissä tietolähteen nimenä käytetään termiä "Web". Kaikkien URL-osoitteiden isäntäosan on sisällettävä kelvollinen verkkotunnuksen nimi (FQDN).
- **Tietokantasyötteet** — On mahdollista luoda mukautettuja liittimiä tietokannan tietueiden syöttämiseksi hakemiston luomista varten.

Kuvassa 4 nähdään hakukoneen syötteiden määrittämisprosessi.

Kuva 4. Haun sisältösyöterajapinta



- 1 Kommentosarja vie tiedot XML-muotoon syötetiedoston luomiseksi
- 2 Kommentosarja lähettää XML-tiedoston hakukoneelle
- 3 Hakukone tuo XML-tiedoston ja indeksoi sisällön

- **Hakukoneen turvallisuus.** Hakukoneen tulee laajentaa yrityksen turvajärjestelmää. Kaikki hakukoneen toiminnot on suoritettava palomuurin takana, ja hakukone saa avata vain muutamia portteja viestinnän mahdollistamiseksi hakukoneen kanssa palomuurin läpi. Hakukone haravoi ja indeksoi sekä julkisia että luottamuksellisia tiedostoja. • Organisaation tiedostotason tietoturvakäytännöt toteutetaan valinnaisella suojauspaketilla.

Oikeiden hakutulosten näyttäminen käyttäjille perustuu haun kohteena olevan hakemiston tulosten suodattamiseen. Sisältö jaetaan kahteen eri luokkaan: julkinen sisältö ja kaikki sisältö. Hakemiston ylläpitäjän määräyksistä riippuen käyttäjät voivat hakea pelkästään hakemiston julkista sisältöä tai sekä julkisia että suojattuja tiedostoja. Oletusarvoisesti sisältö säilyy täysin suojattuna, eikä sitä näytetä käyttäjille, joilla ei ole käyttöoikeuksia.

Hakukoneen on tuettava käyttöoikeuksien perustarkistusta, NTLM (NT LAN Manager) - tarkistusta ja lomakepohjaista käyttöoikeuksien tarkistusta. Hakukone käsittelee kaikkia näitä käyttöoikeuksien tarkistusprotokollia eri tavalla, ja ylläpitäjän on määritettävä ne eri tavalla. Kaikilla edellisillä tavoilla voidaan haravoida ja indeksoida intranetsivustojen suojattuja tiedostoja ja voidaan tarkistaa käyttöoikeudet tiedostoja haettaessa.

Hakukäyttöoikeuksien ohjelmointirajapinnan ansiosta Web-palvelu voi kääntää hakukoneen ohjelmointirajapinnan ja yrityksen palvelimen välillä. Palvelin tarjoaa käyttöoikeuksien hallintapalveluita, joita kutsutaan käyttöoikeusliittimiksi (access connector, AC). Tämä ominaisuus tarjoaa tason hakukoneen ja organisaation käyttöoikeusjärjestelmän välille.

#### Liitinkirjasto

Tämä ohjelmistotasoa muodostaa viesti- ja/tai käyttöoikeustason yrityksen eri tietolähteisiin, kuten rakenteellisiin tietolähteisiin ja yrityssovelluksiin. Tietolähteiden integroimista varten luodaan tavallisesti integrointialusta tai mukautettuja liittimiä tai sovitteita.

Mukautettuja liittimiä voidaan luoda myös seuraavasti:

- Muunna tiedot XML-muotoon hakukoneen syötteessä määritetyssä muodossa.

- Lataa XML hakukoneeseen käyttämällä HTTP-protokollaa.

Taso muodostaa perustan yrityksen tietojen paljastamiseksi lähteen muodosta tai tallennusmuodosta riippumatta. Tason ansiosta palvelut ja eri tietolähteistä luodut integrointisovittimet voidaan siis yhdistää löyhästi.

Tasolla määritetyt metatiedot voivat myös tarjota yrityksen erilaisten lähteiden sisältämien tietojen yhdistämiseen tarvittavan arkkitehtuurin. Yleensä tämä myös mahdollistaa semanttisesti vastaavien kohteiden nopean löytymisen ja noutamisen. Yhdistetyillä metamallilla siis voidaan määrittää yhdenmukaisia ja kontekstin mukaisia metatiedon määritelmiä erilaisten tietojen yhdistämistä varten. Tämä myös parantaa hakutuloksia lisäämällä niihin tietoa asiaan liittyvistä linkeistä sekä jäsennellyistä tieto- ja yrityssovelluksista.

#### Tietolähteet ja -tyypit

Tietoja löytyy nykyään monista eri paikoista yrityksissä. Tietoja on saatavilla jäsennellyssä tai jäsennelemättömässä muodossa sekä omissa tai muiden luomissa tietovarastoissa (ks. taulukko 1). Hakukoneet voivat hakea tietoa, joka voi olla indeksoitua, jolloin se löytyy helpommin ja hakutulokset ovat asiaankuuluvia. On siis ehdottomasti luokiteltava tietolähteet ja käytettävä asiaankuuluvia hakumekanismeja.

Taulukko 1. Tietolähteet ja -tyypit

	PROPRIETARY	NON-PROPRIETARY
Jäsennelly	Tietokantojen hallintajärjestelmät (Oracle, SQL Server, Sybase, Informix)  Tiedostojen hallintajärjestelmät (Documentum, FileNet jne.)  Yrityssovellukset, kuten ERP ja CRM	Tietokantojen hallintajärjestelmät (mySQL)
Jäsennelemätön	Verkon tiedostojärjestelmät/tiedostojärjestelmät  Microsoft Office –asiakirjat  Kuva-, video- ja äänitiedostot	Open office -tiedostot Tiedostojärjestelmät/tiedostopalvelin Tekstitiedostot  Sähköposti

**Jäsennellyt ja jäsentelemättömät tiedot.** Suurin ero jäsennellyjen ja jäsentelemättömien tietojen välillä on ennalta määritelty tietomalli. Jäsennellyllä tiedolla tarkoitetaan sellaista tietoa, joka voidaan laittaa HTML-muodon kaltaiseen ennalta määritettyyn muotoon tai joka sopii taulukkoon, kuten tietokantatekniikassa. Toinen tapa tarkastella jäsennettyä tietoa on metatieto, joka on tietoa tiedosta. Jäsentelemätön tieto puolestaan ei ole ennalta määritellyssä muodossa, joten sitä ei ole liitetty mihinkään. Esimerkkejä tällaisista tiedoista ovat tekstitiedostot ja PDF-tiedostot. Jäsentelemättömässä tietoympäristössä tiedonkeruuseen liittyviä ongelmia aiheuttaa hakutyypin liittyvän työkalun valinta.

#### **Patentilla suojatut ja avoimet tietovarastot.**

Yrityksen tietoja on nykyään saatavilla ERP-, CRM- ja tiedonhallintajärjestelmissä sekä staattisissa ja dynaamisissa portaaleissa. Jotkut näistä järjestelmistä tarjoavat patentilla suojattuja muotoja tietojen julkaisemista varten. Toiset käyttävät avoimia muotoja, kuten XML-muotoa, joka on tasaisesti lisännyt suosiotaan. Monissa sovelluksissa teksti on yhä nykyään pääosassa, mutta myös muiden mediatyyppien, kuten kuvan, äänen ja videon, osuus kasvaa tasaisesti. Tämän vuoksi tarvitaan tehokasta automaattisen tekstihaun, metatietojen (yleensä manuaalisesti luotujen) haun ja sisältöön perustuvan multimediatietojen haun yhdistelmää.

#### **HAKURATKAISUN TARJOAMAT MAHDOLLISUUDET**

Yrityshakuratkaisu voi käynnistää työntekijöiden tuottavuuden kasvamisen ja parempiin tietoihin perustuvan päätöksenteon. Mikä tahansa hakuratkaisu on kuitenkin otettava käyttöön yritys ympäristössä siten, että ymmärretään tarkasti sen arkkitehtuuriset vaatimukset ja toiminnot. Arkkitehtuurin tarkka suunnittelu ja käyttöönotto ovat välttämättä hakuratkaisun tarjoamien merkittävien etujen saavuttamiseksi.

Saat lisätietoja siitä, kuinka ratkaisumme voivat auttaa sinun yritystäsi [ottamalla meihin yhteyttä](#).

#### **MAAILMANLAAJUISTA JOHTAMIS- JA TEKNOLOGIAKONSULTAATIOITA TÄMÄN PÄIVÄN LIIKETOIMINTAYMPÄRISTÖSSÄ**

BearingPoint on maailman johtava maailmanlaajuinen johtamis- ja teknologiakonsultaatioyritys, joka tarjoaa palveluja Global 2000 -yrityksille ja useille maailman suurimmista julkis palveluorganisaatioista. Kokeneet ammattilaisemme auttavat organisaatioita ympäri maailman saavuttamaan tavoitteitaan ja luomaan yritysarvoa. Kohdentamalla asiakkaidemme liiketoimintaprosessit ja tietojärjestelmät autamme heitä saavuttamaan kilpailuetuja — tuottamalla tuloksia nopeasti. Jos haluat lisätietoja, soita numeroon +1 603 589 4089 tai vieraile Web-sivustossamme osoitteessa [www.bearingpoint.com](http://www.bearingpoint.com).

