

## ***Otsing Internetis WWW vahenditega***

---

Info otsimiseks Internetis kasutatakse peamiselt kahte tüüpi veebipõhiseid süsteeme:

- **otsisüsteeme;**
- **temaatilisi katalooge.**

Enamus neist süsteemidest kasutavad oma andmebaasi uute viidete lisamiseks spetsiaalsete otsirobotite abi. Otsirobotid käivad automaatselt WWW-s läbi kõik leitavad lingid ning leides süsteemi andmebaasis veel puuduolevaid lehti, lisavad need sinna. Paraku leiavad sellised süsteemid üles ka alles pooleli olevad ja lihtsalt proovi pärast tehtud lehed ning ka vaatamist mitte väärivad lehed. Sellise automatiseerimise korral pole tagatud ka leitud materjalide korralik struktureerimine.

### **Otsisüsteemid**

---

**Otsisüsteemi ülesanne on Internetist või selle mingist piirkonnast info otsimine etteantud märksõna või fraasi põhjal.**

**OTSIÜSTEEME SAAB LIIGITADA:**

- **teenuste järgi;**
- **otsitava temaatika järgi;**
- **geograafilise piirkonna järgi;**
- **otsitava keele järgi;**
- **fraasi otsimismeetodi järgi.**

**Teenuse järgi** saab eristada peamiselt kolme tüüpi süsteeme:

**WWW otsisüsteemid**, otsivad infot WWW lehekülgedelt.

**Useneti otsisüsteemid**, otsivad infot uudisgruppide, tavaliselt paari viimase kuu materjalide hulgast.

**anonüümse FTP otsisüsteemid**, otsivad materjale FTP serveritest.

**Temaatika järgi** saab eristada:

**Universaalseid otsisüsteeme**, mis otsivad kõikvõimalikke materjale.

**Temaatilisi otsisüsteeme**, mis on keskendunud mingile kindlale teemale (näiteks isikute otsimisele).

**Geograafilise piirkonna järgi** saab eristada:

**Globaalseid otsisüsteeme**, mis hõlmavad kogu Internetti, st maailma.

**Lokaalseid otsisüsteeme**, mis käsitlevad mingis kindlas geograafilises piirkonnas leiduvaid materjale.

Selle viimasega sarnaneb veidi ka otsisüsteemide jaotus **esituskeele** järgi.

**Fraasi otsimismeetodi järgi** eristatakse peamiselt kahte liiki süsteeme:

- **Otsing pealkirja järgi.**
- **Otsing sisu alusel.**

Pealkirja järgi otsimist kasutavad reeglina anonüümse FTP otsisüsteemid, sest sisu vaatamine on komplitseeritud. WWW ja Useneti otsisüsteemid otsivad fraasi harilikult dokumendi seest.

Otsisüsteemid asuvad tänapäeval eranditult WWW lehekülgedel, kuigi ajalooliselt oli see teisiti (näiteks anonüümse FTP tuntuim otsisüsteem *Archie* oli kättesaadav *telneti* vahendusel, praeguseks on kõigile otsisüsteemidele tekkinud oma WWW lehed).

Otsisüsteemile esitatakse harilikult mingi märksõna või fraas, mis iseloomustab otsitava materjali. Vastus kuvatakse WWW leheküljel asuva nimekirjana, pikad nimekirjad mitmel lehel. Iga rida selles nimekirjas viitab mingile leitud materjalile. WWW otsisüsteemid esitavad vastuses lisaks viitele tavaliselt veel ka leitud lehekülje pealkirja ning paar esimest rida.

Leitud materjal seatakse pingeritta märksõnadele vastavuse järgi. Vastavuse hindamiseks on mitmeid erinevaid meetodeid, mis peamiselt vaatavad märksõna esinemist teksti eri osades. Seejuures on märksõna esinemine pealkirjas suurema kaaluga kui kusagil teksti sees või lausa lõpus. Mõned otsisüsteemid väljendavad kokkulangevust isegi protsentides, mida kuvatakse ka nimekirjas.

Harilikult on vaid üks sadade leitud materjalide hulgast meile vajalik. Ülejäänud osa on lihtsalt kasutu ballast. Tuleb meeles pidada, et otsisüsteemid vaatavad ainult teatud sõnade esinemist, mitte ei mõista teksti sisu. Ometigi on inimese jaoks saja või paarisaja leitud lehekülje läbivaatamine lihtsam kui kogu WWW läbiuurimine.

Erijuhul võib ballast hoopiski puududa, näiteks kui teame mingi kindla faili nime, mille FTP otsisüsteem kergesti leiab.

## **Temaatilised kataloogid**

---

Temaatilised kataloogid kujutavad endist hierarhilisi katalooge, mis järjest peenemalt käsitletavat materjali liigitavad. Nii on valdkonna “teadus” üks alamvaldkond “bioloogia”, selle alamvaldkond “merebioloogia” jne. Iga alamvaldkond võib sisaldada hulga viiteid konkreetsetele materjalidele. Temaatilises kataloogis asendub otsimine navigeerimisega, milles kitsendatakse järk-järgult otsitavat valdkonda.

Temaatilised kataloogid on reeglina olemas vaid WWW-s esitatud materjali kohta. Uudisegruppides esitatav materjal on juba eelnevalt eri valdkondadesse e. gruppidesse jaotatud, anonüümse FTP puhul on aga materjalidest näha vaid faili nimi, mille järgi on liigitamine pea-aegu võimatu.

Temaatilisi katalooge saab liigitada samuti nagu otsisüsteeme:

- temaatika järgi;
- geograafilise piirkonna järgi;
- kasutatava keele järgi.

Väga palju temaatilisi katalooge on ühitatud vastava WWW otsisüsteemiga. Ka eri liiki otsingud (näiteks WWW-s ja Usenetis) võivad konkreetsetes otsisüsteemides ühitatud olla. Nii on Internetis palju liitsüsteeme, mis katavad suure osa käsitletud valdkondadest, kuid ükski neist pole kõikvõimas.