

Wikidata – mis, miks ja kuidas?

Oled Sa kunagi mõelnud järgmistele küsimustele?

- Mis on enim levinud perekonnanimed Eestis?
- Kelle sünnipäev täna on?
- Millised lennujaamad asuvad Tallinnast 100 kilomeetri raadiuses?

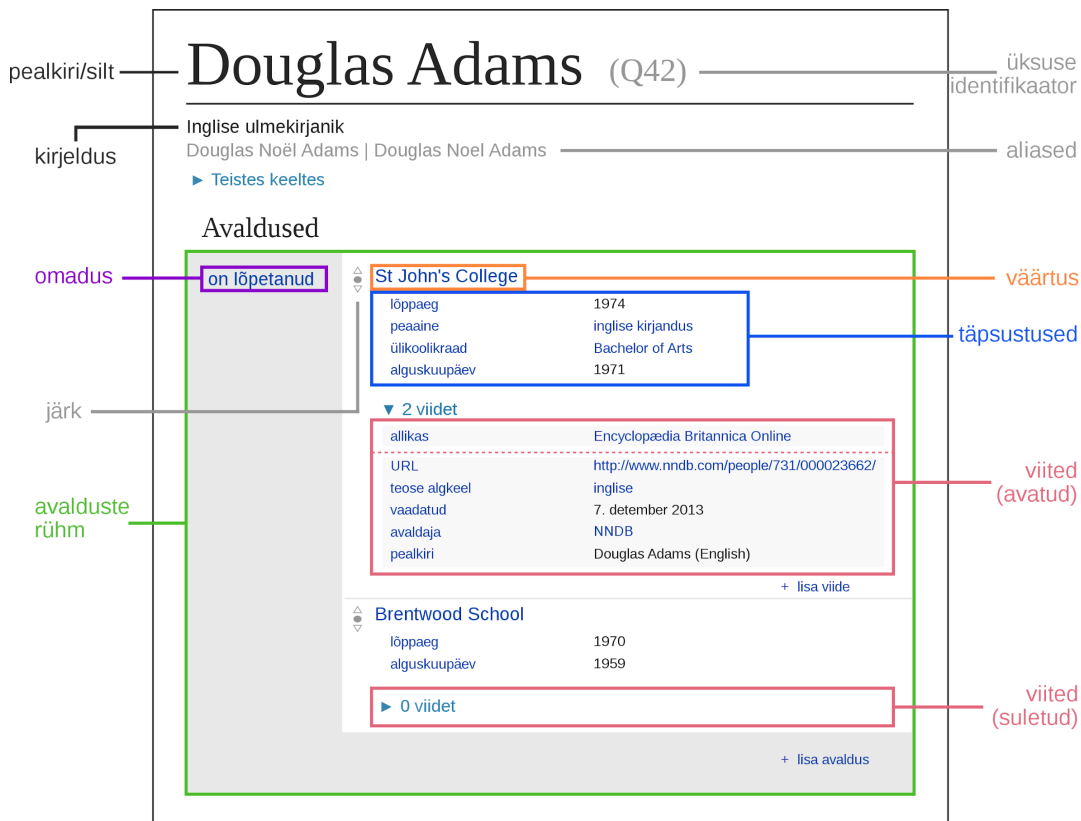
Wikidatas on vastused olemas.

Mis on Wikidata?

[Wikidata](#) ehk Vikiaandmed on vaba koondandmebaas, mida saavad lugeda ja täiendada nii inimesed kui masinad. Teadmuspank toimetab vabatahtlike koostöös ja on loodud Vikipeedia ja sõsarprojektide ühiseks andmeallikaks. Kuna andmed on Wikidatas avaldatud [Creative Commons avaliku omandi litsentsi](#) alusel, siis võib neid igaüks muuta, jagada ja rakendada, seda soovi korral ka ärilisel eesmärgil. Erinevaid andmetüüpe on palju: tekst, pildid, arvud, koordinaadid, kuupäevad jpm.

Andmete ülesehitus

Andmed Wikidatas on seotud ja struktureeritud. See tähendab, et Wikidatas ei ole mitte ainult sissekanded, vaid ka nende allikad ja seosed teiste andmebaasidega. Tänu sellele on igal inimesel võimalik andmeid omaenda projektis hõlpsalt kasutada.



The screenshot shows the Wikidata page for Douglas Adams (Q42). The page is annotated with labels explaining its structure:

- pealkiri/silt**: Douglas Adams (Q42)
- üksuse identifikaator**: (Q42)
- kirjeldus**: Inglise ulmekirjanik
- aliasid**: Douglas Noél Adams | Douglas Noel Adams
- omadus**: on lõpetanud
- järk**: St John's College
- avalduste rühm**: avaldused
- väärtus**: 1974
- täpsustused**: inglise kirjandus
- viited (avatud)**: Encyclopædia Britannica Online
- viited (suletud)**: 0 viidet

The page also shows a table of publications (Avaldused) for Douglas Adams, including details like the source (allikas), URL, language (teose algkeel), date (vaadatud), publisher (avaldaja), and title (pealkiri).

Andmeüksus ulmekirjanik Douglas Adamsi kirje näitel.

Andmeid esitatakse Wikidatas andmeüksustena (ingl *item*). Üksus võib olla konkreetne objekt või isik, liiginimi, mõiste või midagi muud. Igal üksusel on unikaalne identifikaator ehk kood kujul Q1234567, mille eesmärk on siduda samad nähtused eri keeltes ning võimaldada masinatel andmeid kergemini lugeda.

Iga üksuse juurde kuuluvad avaldused (ingl *statement*) ehk andmete sissekanded. Avaldusi väljendatakse omaduste (ingl *property*) ja väärtuste (ingl *value*) paaridena. Väärtused võivad olla arvandmed, pildid, aga ka teised Wikidata andmeüksused. Näiteks kirjanik Douglas Adams on lõpetanud (omadus) *St John's College*'i (väärtus).

Ka omadustel ja väärtustel on unikaalsed identifikaatorid kujul Q1234567. Parema arusaama saamiseks Wikidata ülesehitusest külasta [Wikidata tuure](#).

Miks Wikidatat vaja on?

Wikidatal on oluline roll andmetepõhise interneti arengus. Vikipeedia kasutab Wikidatat, et siduda omavahel sama artikli versioonid erinevates keeltes. Wikidata võimaldab ka Vikipeedia infokastides andmeid automaatselt uuendada.

Wikidatat kasutavad ka ettevõtted. Google'i teadmuspank Knowledge Graph kasutab Wikidatat paremate otsingutulemuste kuvamiseks infokastides (vt pilt paremal).

Lisaks on väga palju veebilehti, mis toimetavad Wikidata baasil. [Wikidata Guessr](#) on asukoha ära arvamise mäng, [Histropedia](#) võimaldab luua ajatelgi, [Monumental](#) näitab kultuuripärandit, [Scholia](#) teaduslikku infot, [Ask Platypus](#) vastab küsimustele, [Crotos](#) on kunstikogu... Pikem nimekiri on [siin](#).

Kuidas minul Wikidatast kasu on?

Wikidatas on vastused neile küsimustele, millele internetiotsing vastata ei oska. See võimaldab andmeid kergelt analüüsida ja visualiseerida. Sa võid vähese vaevaga luua omaenda äpi või veebilehe, mis kasutab vikiandmeid.

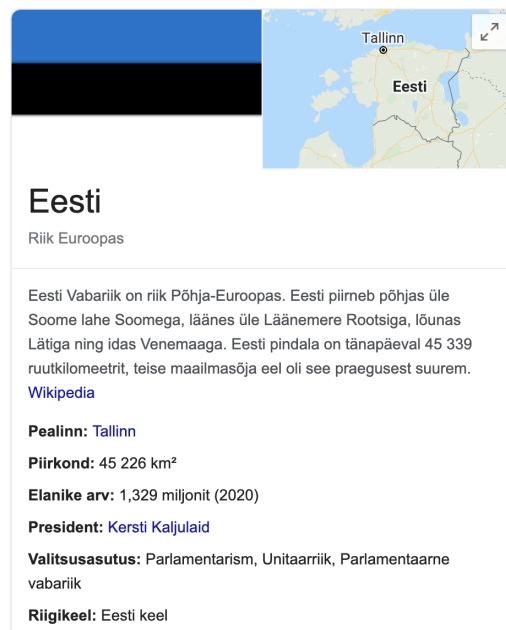
Kuidas ma Wikidatast midagi küsin?

Kuna Wikidata on andmebaas, tuleb küsimusi esitada masinsõbralikul kujul, kasutades andmepäringuid. Päringuid kirjutame keeles SPARQL, aga ära muretse, kui see kõlab liiga tehniliselt – selles juhendis vaatame lihtsaid näidispäringuid ja kuidas neid muuta.

Andmepäringute koostamine

Maailma tuntud kassid

Vaatame lihtsat näidet: tahame näha nimekirja kõigist tuntud kassidest maailmas. Päringute tegemiseks kasutame [Wikidata päringuteenust](#).



Eesti
Riik Euroopas

Eesti Vabariik on riik Põhja-Euroopas. Eesti piirneb põhjas üle Soome lahe Soomega, läänes üle Läänemere Rootsi, lõunas Lätiga ning idas Venemaaga. Eesti pindala on tänapäeval 45 339 ruutkilomeetrit, teise maailmasõja eel oli see praegusest suurem.

[Wikipedia](#)

Pealinn: [Tallinn](#)
Piirkond: 45 226 km²
Elanike arv: 1,329 miljonit (2020)
President: [Kersti Kaljulaid](#)
Valitsusasutus: Parlamentarism, Unitaarriik, Parlamentaarne vabariik
Riigikeel: Eesti keel

Kui Su veebileht on inglise keeles, klikki ülal paremal English ja kirjuta sisse eesti.

1. Klikki *Näited*
2. Vali rippmenüüst *Cats*
3. Vajuta vasakul sinisele  nupule

Nüüd on meil nimekiri kõigist kuulsatest kassidest internetis – või vähemalt nendest, kellest Wikidata teadlik on.

Mida kõik need sõnad tähendavad?

Sinu ees on SPARQL-is kirjutatud päring. Vaatame lähemalt, mida iga sõna tähendab ja mis rolli ta mängib.

```
#Cats
SELECT ?item ?itemLabel
WHERE
{
  ?item wdt:P31 wd:Q146.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

- **#Cats** Kõik, mis algab trellidega (#), on **kommentaär**. Seda arvuti ignoreerib.
- **SELECT** Iga päring **valib** midagi, mida hiljem **kuvada**. Mis andmed päriselt ekraanile jõuavad, kirjeldatakse hiljem. Selle rea eesmärk on anda arvutile mingid muutujad (praegu *?item* ja *?itemLabel*), et neid hiljem tulemustes näidata.
- **?item** on suvaline **muutujanimi**, sama hästi võiks olla *?loom* või midagi muud. Muutuja tunned ära küsimärgi järgi. Selle muutuja külge seotakse nimekiri meie soovitud tulemustega.
- **?itemLabel** on samuti muutuja, mis võimaldab päringu tulemusena näha silte ehk **üksuste nimetusi inimkeeles**. Tänu sellele muutujale näidatakse näiteks üksuse Q378619 juures meile vastava kassi silti (ehk nime) “CC”.
- **WHERE** Siin me kirjeldame, **mis väärtused muutujatele antakse**. WHERE-lause algab ja lõpeb loogeliste sulgudega { }.
- **wdt:P31** **wdt:** on **omaduse eesliide** ning **P31** on konkreetne omadus *üksikjuht nähtusest* (näed seda ka hiirega omaduse peale liikudes). Erinevaid omadusi on väga palju, näiteks *P106* (amet) või *P1476* (pealkiri). Vaata täispikka nimekirja [siit](#).
- **wd:Q146** **wd:** on **väärtuse eesliide** ning **Q146** on konkreetne väärtus *kass*. Pane tähele, et omadused algavad P-tähega ning väärtused Q-ga. Omaduse ja väärtuse paar annavad praegu kokku *üksikjuht nähtusest kass*. Punkt rea lõpus annab arvutile mõista, et lause sai läbi.
- **SERVICE** See maagiline rida ütleb arvutile, et tahame **kuvada inimkeelseid silte**, praegusel juhul täpsemalt ingliskeelseid. Praegu pole oluline seda rida mõista.

Päris palju, mida seedida. Võib-olla aitab, kui selle päringu inimkeelde tõlgime: “Anna mulle nimekiri üksustest koos siltidega, kus iga üksus on *üksikjuht nähtusest kass*. Seejärel loo üksustele ingliskeelsed sildid.”

Tulemuseks saime kaks tulpa: *item* ja *itemLabel* (muutujanime järgi). Vasakus tulpas on üksuste masinloetavad koodid ja paremas nende ingliskeelsed sildid. Kui klõpsata vasakus tulpas mõnel koodil ehk identifikaatoril, avaneb selle üksuse Wikidata lehekülg. Selleks, et tulemusi sorteerida, klikki vastava tulba nimel (nt vajutades *itemLabel* tulpale sorteeritakse tulemused tähestiku järjekorras).

Aga koerad?

Nagu mainitud, on igal üksusel Wikidatas unikaalne identifikaator ehk lihtsamalt öeldes oma kood. Vaatame kassinäidet ja proovime seda muuta, et kuvada hoopis koeri.

1. Kustuta tekstikastist *Q146* (jätta alles *wd:*) ja kirjuta selle asemele *koer*
2. Vajuta Ctrl+Tühik (või Ctrl+Alt+Tühik või Alt+Enter)
3. Enterit vajutades asendatakse *koer* koodiga *Q144*

(Võid kassi koodi *Q146* asemele ka kohe *Q144* kirjutada. See on eeldusel, et tead peast koera identifikaatorit Wikidatas!)

Aga Eesti?

Vaatame Eesti luuletajaid. Ava [järgmine näidispäring](#) ja pane tööle. Päring kuvab kõik Eesti luuletajad, kellel on Wikidatas ka sünnikoht täpsustatud. Tulemus on aga visuaalselt igav: palju parem oleks seda kaardil vaadata. Proovime teha muudatusi, et seda saavutada.

Kõigepealt proovime aru saada, kuidas praegune päring töötab. Meil on kaks muutujat: *?üksus* ja *?sünnikoht*. Esmalt täpsustame, et tahame muutujasse *?üksus* inimesi, seejärel lisame, et nad peavad olema ka Eesti kodanikud¹ ja ametilt luuletajad. Viimaseks nõuame, et neil oleks sünnikoht täpsustatud. Selle asemel, et öelda, et sünnikoht peab olema kindlasti Tallinn või Saaremaa, määrame sünnikoha hoiustamiseks uue muutuja *?sünnikoht* – meil ei ole vaja piirata, kus ta täpselt sündis, vaid tahame, et sünnikoha andmed oleksid olemas.

Nüüd aga tahame kuidagi sünnikohti kaardil näidata. Selleks on meil vaja koordinaate. Lisame uue rea, mis nõuab, et sünnikohal (just sünnikohal, mitte üksusel) oleks olemas koordinaadid – ja salvestame need koordinaadid uude muutujasse, näiteks *?koord*. Märkad sarnasust eelmise reaga? Kuna tahame koordinaate ka ekraanil kuvada, siis peame SELECT reale koordinaate hoidva muutuja juurde lisama. Proovi esmalt ise, vajadusel vaata [minu lahendust](#). (Nipp: on olemas omadus *koordinaadid* koodiga *P625*.)

Nüüd on meil koordinaadid, aga kaarti endiselt mitte! Kuvamisviisi muutmiseks on all vasakul servas silmaikoon. Klikki sellel ja vali *Map*. Ongi olemas! Võid testida erinevaid esitusviise – enamik neist on selle päringu tulemustega üsna kummalised.

¹ Märkus: paljudel inimestel ei ole kodakondsuse parameeter (veel) märgitud.



```
8 ?üksus wdt:P19 ?sünnikoht. # kellel on täpsustatud sünnikoht, mille jaoks me tahame uut muutujat
9 ?sünnikoht wdt:P625 ?koord. # sünnikohal on koordinaadid, mille jaoks me tahame uut muutujat
10
11 SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
12 }
```

üksus	üksusLabel	sünnikohtLabel
wd:Q671861	Toomas Liiv	Tallinn

Oluline järelendus: selleks, et päringut koostada, ei pea me iga kord ratast leiutama ega SPARQL-i suurepäraselt oskama. Piisab sellest, kui me oskame mõnd teist päringut enda soovi järgi kohendada.

Proovi ise!

Proovi nüüd täiesti iseseisvalt päring koostada. Leia kõik eestlased, kes on poliitikud ja ühtlasi lauljad. Abiks võib tulla eelnevalt koostatud päring...

Kui jääd jänni või tahad tulemust kontrollida, vaata [minu lahendust](#).

Kas ma oskan nüüd ise päringut koostada?

Jah! Näidete muutmise teel on võimalik väga palju ära teha. Kustuta ridu, lisa juurde, kombineeri erinevaid päringuid. Nipp: kui tahad sõbraga enda koostatud päringut jagada, vali vasakul menüüst ketikujuline nupuke. See tekitab veebilingi käesoleva päringuga.

Kui SPARQL hakkas huvi pakkuma ja tahaksid päringute koostamise kohta edasi õppida, külasta [seda juhendit](#).

Mida ma veel teha saan?

Siin on pikk nimekiri juba koostatud päringutest, mida saad katsetada ja muuta. Lisaks on väga palju näidispäringuid olemas [päringukeskkonna](#) *Näidete* kastis (kust me kasside näidet juba vaatasime).

Kui tahad Wikidatasse panustada, on võimalusi palju. Õpi [tuuridel](#) üksusi lisama ja muutma või hüppa kohe vette ja täienda [juhuslikku kirjet](#). Lisa juurde andmeid, näiteks pilte [Eesti kultuurimälestistest](#). [Tõlgi keskkonda](#) eesti keelde või [anneta Wikidatasse andmeid](#).

Programmeerijale: samal real, kus muutsime kuvamisviisi, on paremal ka nupud tulemuse allalaadimiseks masinloetaval kujul ning päringu nägemiseks eri programmeerimiskeeltes.

Näidispäringuid

Vastused sissejuhatuses püstitatud küsimustele

- [Lennujaamad, mis asuvad Tallinnast 100 km raadiuses](#)
- [Mis on enim levinud perekonnanimed Eestis?](#)
- [Kelle sünnipäev täna on?](#)

Tartu Kunstimuuseum

- [Nikolai Triigi teosed ajajoonel](#)
- [TARTMUS maalikogu: kas autori looming kuulub avalikku omandisse ning mitu selle autori teost on kogus](#)
- [Maalid, mis kujutavad Eestis voolavaid jõgesid](#)
- [Millistel TARTMUS maalidel on kujutatud mõnda veekogu](#)
- [Millised kümme TARTMUSile kuuluvat maali on kõige suurema pindalaga](#)
- [Millal on TARTMUS maalikogusse kuuluvad teosed loodud](#)
- [Kui pikad on TARTMUS kogusse kuuluvate maalide nimed](#)
- [TARTMUS kogus olevad teosed, mis kujutavad konkreetset inimest ja teave kujutatu soo kohta](#)

Veel kunsti

- [Kõik maalid, mis kujutavad Eestit või midagi Eestis](#)
- [Kõik maalid, mis kujutavad Soomet või midagi Soomes](#)

Poeetid

- [Eesti poeetid](#)
- [Poeetid, kes kõnelevad/kirjutavad eesti keeles](#)
- [Eesti poeetid ja nende sünnikohad](#)
- [Eesti poeetid, kes sündisid pärast 1950. aastat](#)

Eestlased

- [Kõik eestlased](#)
- [Eestlased, kes sündisid kas Tartus või Tallinnas](#)
- [Kõik eestlased, välja arvatud need, kes sündisid Tallinnas](#)
- [Eestlased, kelle nimi on Jaan](#)
- [Eesti poliitikud, kelle isa oli/on samuti poliitik](#)
- [Levinud eesti eesnimed](#)
- [Isikuartiklite jaotus Eesti Vikipeedias \(top 50 ametit\)](#)
- [Isikuartiklite jaotus Eesti Vikipeedias \(top 50 ametit\) – ainult naised](#)
- [Eesti emad, kellel on vähemalt 3 last](#)
- [\(Endised\) ministrid, kelle ema või isa on samuti olnud minister](#)
- [Eesti peaministrid, järjestatud ametisoldud aja järgi](#)

Veel Eesti kohta

- [Eesti asjad kaardil](#)
- [Eesti järved kaardil](#)
- [Lahingud, mis on toimunud Balti riikides](#)
- [Koolid, mis asutati enne 1900. aastat \(sh endised\)](#)
- [Kõik Eesti kõrgkoolid \(sh endised\)](#)

Muu

- [Inimesed, kellel on diagnoositud COVID-19](#)
- [Tšingis-khaani järeltulijad](#)
- [Top 100 inimeste surmapõhjust](#)
- [Loomad, kellel on poliitiline ametikoht](#)



DIGIKULTUURI
AASTA
2020