Annif

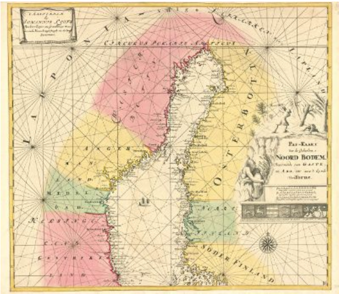
automaattinen asiasanoitus ja luokitus kirjan tuotetietojen ja sisällön perusteella

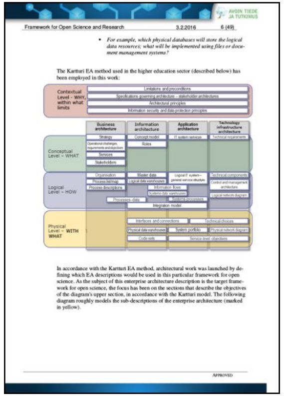
Osma Suominen

Kirja-alan metatietoseminaari 20.3.2019

Extrablad till
ÅBO UNDERRÄTTELSER
N:o 2.
Finlands oavhängighet.
Lantdagen omfattar regeringens proklamation om Finlands fullständiga oavhängighet och ansluter sig till huvudprinciperna i regeringens program för tryggandet av landets nya ställning. Beslutet fattades med 100 röster emot 88 vilka tillföllo ett av socialdemokraterna formulerat förslag.
Våldsdåd av arbetare.
Återupptagandet av trafiken Åbo—Stockholm.
Pris 25 penni.

VENNY SOLDAN-BROFELDT.
Unikko. (Papaver orientale.)




Framework for Open Science and Research 3.2.2016 6 (49)

- *For example, which physical databases will store the logical data resources; what will be implemented using files or document management systems?*

The Katturi EA method used in the higher education sector (described below) has been employed in this work:

In accordance with the Katturi EA method, architectural work was launched by defining which EA descriptions would be used in this particular framework for open science. As the subject of this enterprise architecture description is the target framework for open science, the focus has been on the sections that describe the objectives of the diagram's upper section, in accordance with the Katturi model. The following diagram roughly models the sub-descriptions of the enterprise architecture (marked in yellow).

APPROVED

Extrablad till

ÅBO UNDERRÄTTELSER

N:o 2.

Åbo, den 6 december 1917.

Finlands oavhängighet.

Lantdagen omfattar regeringens proklamation om Finlands fullständiga oavhängighet och ansluter sig till huvudprinciperna i regeringens program för trygggandet av landets nya ställning. Beslutet fattades med 100 röster mot 88 vilka tillföllo ett av socialdemokraterna formuleat förslag.

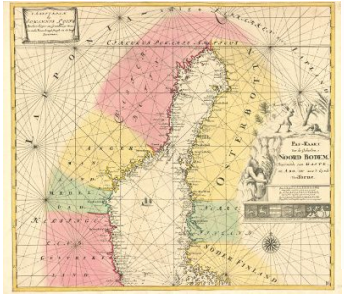
Vår speciella korrespondent från i Helsingfors meddelar:

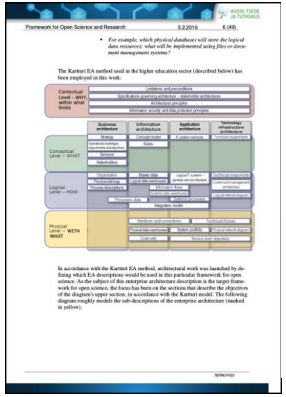
Den 6:nde blåste vi gjorde i dag c. 2 uppmärksamma att en de hemmiljöpromenaj beneddlil Finlands till att Finlands lantdag skulle ha en sammanträdande den sista oktober av nuvarande underhandlingsskedet.

Förslagen lyda:

Med tillämpning av regeringen den 4 December 1917 av regeringen lämna, offrar de Finland ad dipl. avsnitt Finlands oavhängighet under förutsättning att Ryska ministerns underhandlingar med representanter för Finlands folkrepresentation ej leopard i sitt sätt Finland hemlig Ryska Ryssland Finlands Finnlands, genom att proklamera Finlands oavhängighet, i det de fullföra Finlands ställning i sitt sätt som en av de stater som i sitt sätt som en av de stater som en fri, suveränaste enhetsstat. Ett betecknings av utrikesfrågor samt representanter av Ryska riksen att i

Pris 25 penni.





Framework for Open Science and Research 3.2.2016 6 (49)

- *For example, which physical databases will store the logical data resources; what will be implemented using files or document management systems?*

The Kartturi EA method used in the higher education sector (described below) has been employed in this work:

In accordance with the Kartturi EA method, architectural work was launched by defining which EA descriptions would be used in this particular framework for open science. As the subject of this enterprise architecture description is the target framework for open science, the focus has been on the sections that describe the objectives of the diagram's upper section, in accordance with the Kartturi model. The following diagram roughly models the sub-descriptions of the enterprise architecture (marked in yellow).

APPROVED

Extrablad till

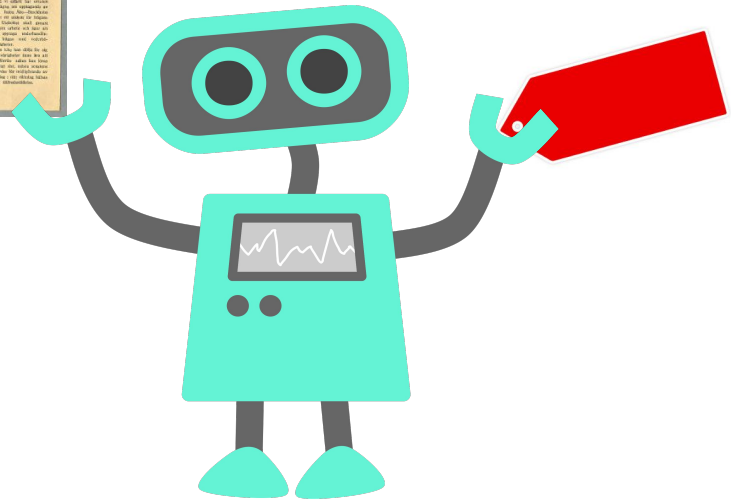
ÅBO UNDERRÄTTELSER

N:o 2.

Finlands oavhängighet.

Lantdagen omfattar regeringens proklamation om Finlands fullständiga oavhängighet och ansluter sig till huvudprinciperna i regeringens program för tryggandet av landets nya ställning. Beslutet fattades med 100 röster emot 88 vilka tillföllo ett av socialdemokraterna formulerat förslag.

Pris 25 penni.



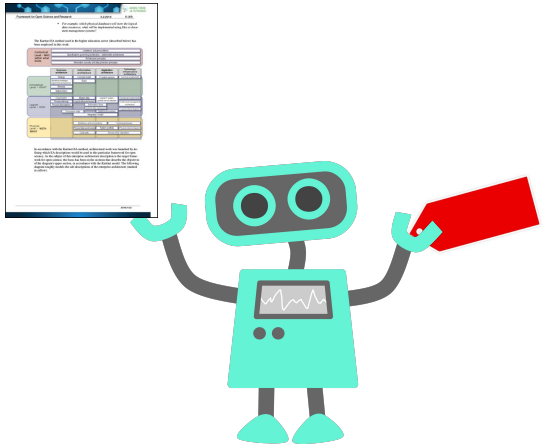
Annif

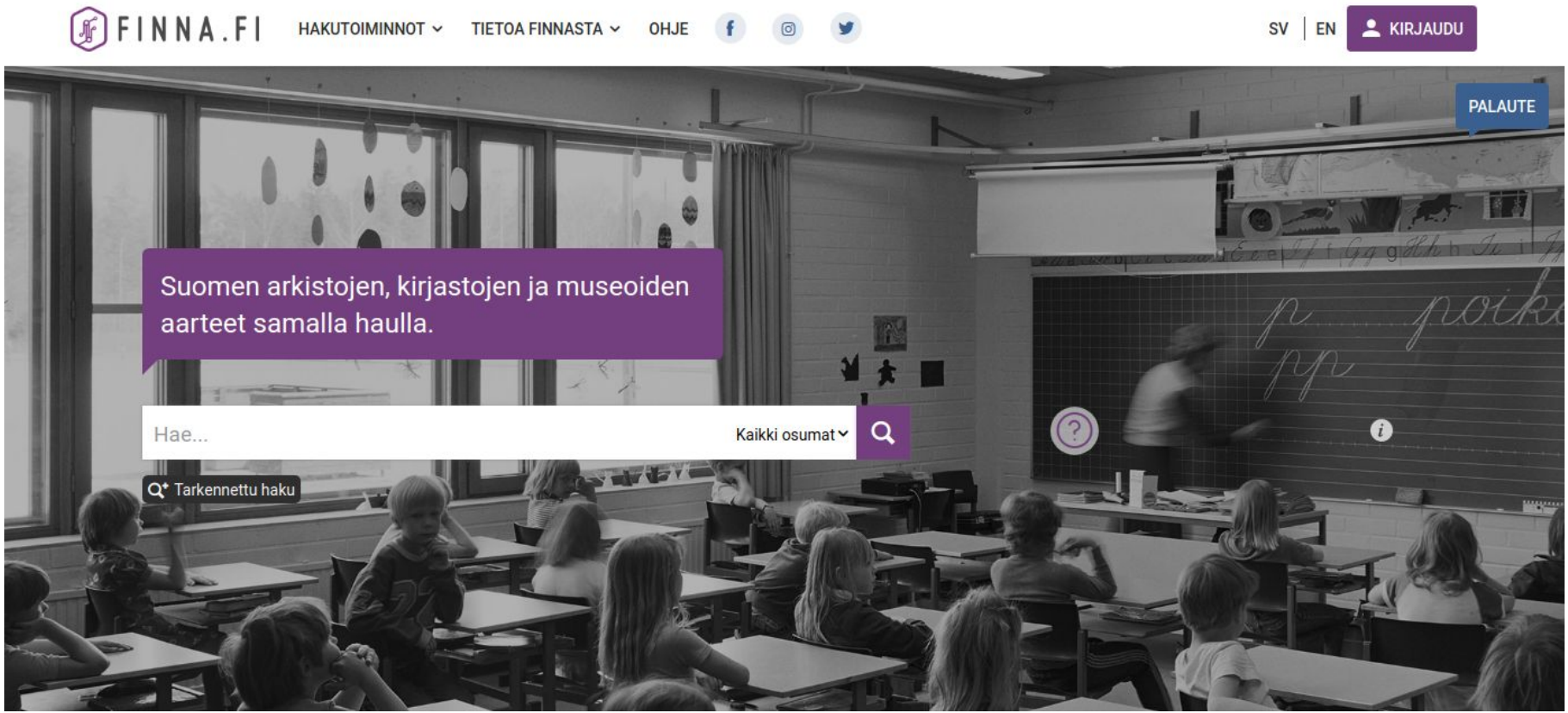
Tuottaa sisällönkuvailua (asiasanoitus ja luokitus) tekstin perusteella

Kansalliskirjastossa kehitetty työkalu, avointa koodia

Soveltuu parhaiten asiatekstille, mutta testataan myös muilla aineistoilla

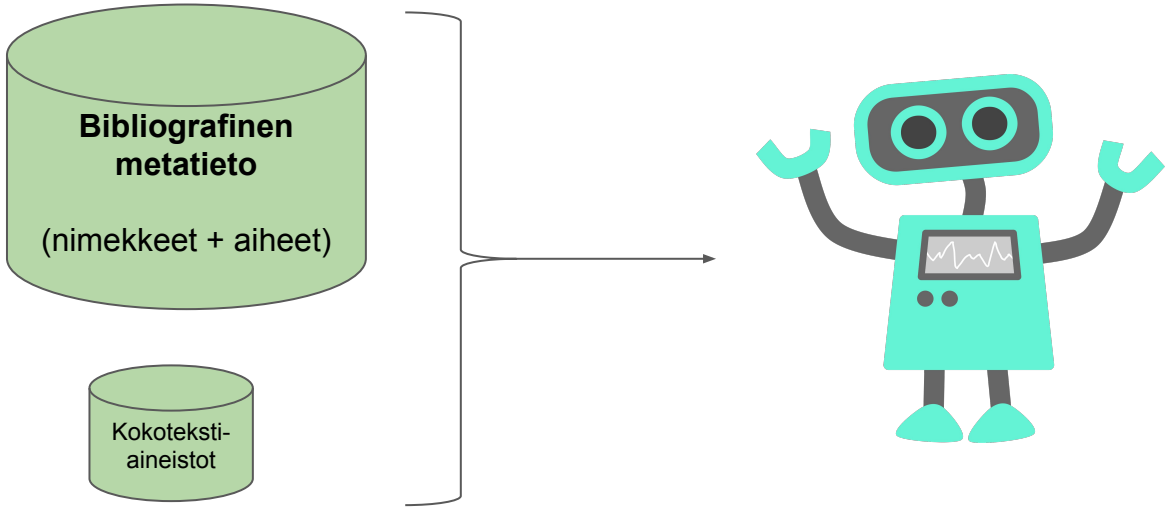
Perustuu **kieliteknologiaan** ja **koneoppimiseen**





Paljon muistiorganisaatioiden metatietoa, mm. **15M tietuetta** Finnassa

Annif oppii aineistosta, miten tunnistaa aiheita



Esimerkkituloksia esittelyteksteillä

Lomake testaamiseen: http://annif.org

Asumiskelvoton maapallo - Elämä lämpenemisen jälkeen

Otava, ISBN: 9789511349358
David Wallace-Wells, kääntäjä Tarja Kontro
Kustantajan asiasanat:
ekokatastrofit; ilmasto; ilmastonmuutokset; kasvihuoneilmiö; lämpeneminen; maapallo; tulevaisuus; uhkakuvat; ympäristö

Text to analyze:

Asumiskelvoton maapallo - Elämä lämpenemisen jälkeen

Kaikki mitä sinun pitää tietää ilmaston lämpenemisestä

Tilanne on pahempi kuin tiedämme. Ilmastonmuutoksesta käytävässä keskustelussa vähätellään ilmiön vakavuutta. Kirja kuvaa vakuuttavasti tieteelliseen tutkimukseen nojautuen ilmastonkriisin vaikutuksia eri elämänalueilla. Samalla kirja kutsuu ihmisiä toimimaan - kun vielä voimme.

Kansanvälinen menestysteos.

Project (vocabulary and language):

YSO ensemble Finnish

Analyze

Results

- lämpeneminen
- ilmastonmuutokset
- ilmasto
- kasvihuoneilmiö
- kirjat
- paha
- käytävät
- maapallo
- ilmiöt
- ilmakehä

Hjallis Harkimo. Bisneselämäkerta

WSOY, ISBN: 9789510443613

Tommi Aitio

Text to analyze:

Hjallis Harkimo. Bisneselämäkerta

Kuinka tienasin miljoonani ja melkein menetin kaiken.
Hjallis Harkimon bisneselämäkerta paljastaa, mistä menestynyt liikemies on tehty. Sitä lukee kuin dekkaria, ja se tarjoaa käytännöllisiä vinkkejä menestyvän bisneksen rakentamiseen.
Harry ”Hjallis” Harkimon epäsuomalainen, häpeämättömän suorasukainen tyyli on ihastuttanut ja vihastuttanut jo yli kolmenkymmenen vuoden ajan. Harkimo nousi koko kansan tietoisuuteen 1980-luvulla purjehtiessaan yksin maailman ympäri, ensimmäisenä suomalaisena. Kisassa kolmannen sijan tuonut purjehdus oli myös äärimmäinen bisnesoperaatio, joka oli viedä Harkimon hengen.
Tuosta lähtien Hjallis on ollut mukana monissa erilaisissa liiketoimissa, joista tunnetuimmat ovat jääkiekkojoukkue Jokerit, erilaiset urheilu- ja monitoimihallit, Sipoonrannan asuinalue sekä monien huippu-urheilijoiden managerointi. Hjalliksen näkyvästä toiminnasta huolimatta monille on jäänyt epäselväksi, mistä miehen rahat oikein tulevat. Kirjassaan Harkimo paljastaa liiketoimiensa taustat sekä metodinsa bisneksen tekemiseen.

Project (vocabulary and language):

YSO ensemble Finnish

Analyze

Results

- urheilijat
- tietoisuus
- urheilu
- miehet
- monitoimihallit
- Hartwall Areena
- asuinalueet
- liikemiehet
- salapoliisikirjallisuus
- purjehdus

Rooman kuningas

Like, ISBN: 9789520118662

Francesco Totti ja Paolo Condò, kääntäjä Juuso Kortelainen

Text to analyze:

Rooman kuningas

Jalkapalloilija Italian ja maailman huipulta

Perusteellinen omaelämäkerta jalkapalloilijasta, joka rakastaa lajiaan ja seuraansa varauksetta, intohimoisesti ja ikuisesti - ja jota fanit kautta maailman palvovat samalla palolla. Elämää kentällä ja sen ulkopuolella pikku pallopojasta lajin ehdottomaan kärkeen niin Italiassa kuin koko maailmassa. Ura, joka ei edennyt kuin unelma, vaikka sen lopulta täyttikin.

Project (vocabulary and language):

YSO ensemble Finnish

Analyze

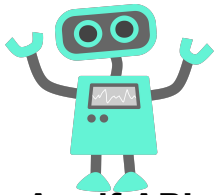
Results

- jalkapalloilijat
- ihailijat
- italian kieli
- kuninkaat
- unelmat
- varaukset
- muistelmat
- lajit
- maailma
- pallot

Annif API esimerkki

“The quick brown fox jumped over the lazy dog.”

Analyze this!



Annif API

results=[
{uri=”<http://www.yso.fi/onto/yso/p2228>”, score=0.2595, label=”**red fox**”},
{uri=”<http://www.yso.fi/onto/yso/p5319>”, score=0.2039, label=”**dog**”},
{uri=”<http://www.yso.fi/onto/yso/p8122>”, score=0.1946, label=”**laziness**”},
{uri=”<http://www.yso.fi/onto/yso/p25726>”, score=0.1285, label=”**brown**”},
{uri=”<http://www.yso.fi/onto/yso/p4760>”, score=0.1220, label=”**triple jump**”}
]

Tulevaisuus

Inari Haapaniemi, Kirjavälitys
Osma Suominen, Kansalliskirjasto

Kirjan elinkaari – mitkä tiedot muuttuvat

Ennen ilmestymistä | Ilmestymisen jälkeen

Kevään/kesän/syksyn katalogit ja listat

Ennakkotieto vain kustantajalla

Julkinen ennakkotieto

- Jopa 6-7 kuukautta ennen kirjan julkaisupäivää ☐ Lähes kaikki tiedot voivat muuttua
- Ennakkotieto toimitetaan
  - jälleenmyyjille
  - Kansalliskirjastolle
  - KV:n sopimuskirjastoille
- Riski, että vanhentuneita tietoja jää elämään, mikäli vastaanottajan järjestelmät eivät käsittele kaikkea päivitystietoa.

Tarkentuva ennakkotieto

- Lopulliset kansikuvat ja esittelytekstit
- Ilmestymispäivät
- Tekijätiedot ja luokittelut
- Mittatiedot ja sivumäärät

Tarkistettu tieto

- Metadataa ja tuotetta verrattu keskenään ja poikkeamat tiedoissa korjattu
  - KV
  - Kansalliskirjasto - Fennica
- Yleensä muuttuvat enää
  - painostiedot
  - saatavuus / varastosaldot
  - hinnat silloin tällöin

ONIX-formaatissa muutoksia
- uusien koodien käyttöönotto
- suositukset koodien käytöstä muuttuvat

☐ päivitykset takautuvasti

Kirjan elinkaari - tulevaisuus?

Ennen ilmestymistä | Ilmestymisen jälkeen

Kevään/kesän/syksyn katalogit ja listat

Ennakkotieto vain kustantajalla

Julkinen ennakkotieto

- Jopa 6-7 kuukautta ennen kirjan julkaisupäivää ☐ Lähes kaikki tiedot voivat muuttua.
- Ennakkotieto toimitetaan
  - jälleenmyyjille
  - Kansalliskirjastolle
  - KV:n sopimuskirjastoille
- Riski, että vanhentuneita tietoja jää elämään, mikäli vastaanottajan järjestelmät eivät käsittele kaikkea päivitystietoa.

Tarkentuva ennakkotieto

- Lopulliset kansikuvat ja esittelytekstit
- Ilmestymispäivät
- Tekijätiedot ja luokittelut
- Mittatiedot ja sivumäärät

”Sisältötarkistettu” ennakkotieto

”Painovalmiin” aineiston perusteella
- luokittelut
- asiasanoitus

automaation avulla

- TIETO KIRJASTOJEN JA KAUPPOJEN KÄYTETTÄVISSÄ ENNEN KIRJAN ILMESTYMISTÄ
- TÄLLÄ HETKELLÄ TIETO TULEE JÄLKIJÄTTÖISESTI
- OSA FENNICA-LUETTELOINNISTA?

Tarkistettu tieto

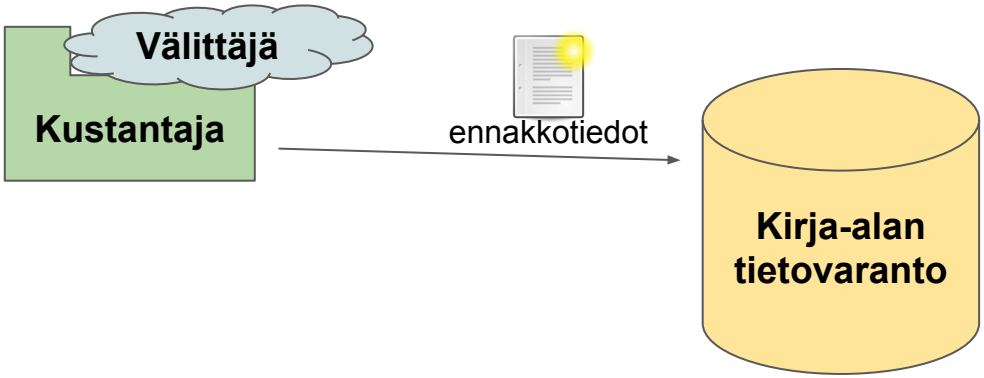
- metadataa ja tuotetta verrattu keskenään ja poikkeamat tiedoissa korjattu
- yleensä muuttuvat
  - hinnat
  - saatavuus / varastosaldot
  - painostiedot

ONIX-formaatissa muutoksia
- uusien koodien käyttöönotto
- suositukset koodien käytöstä muuttuvat

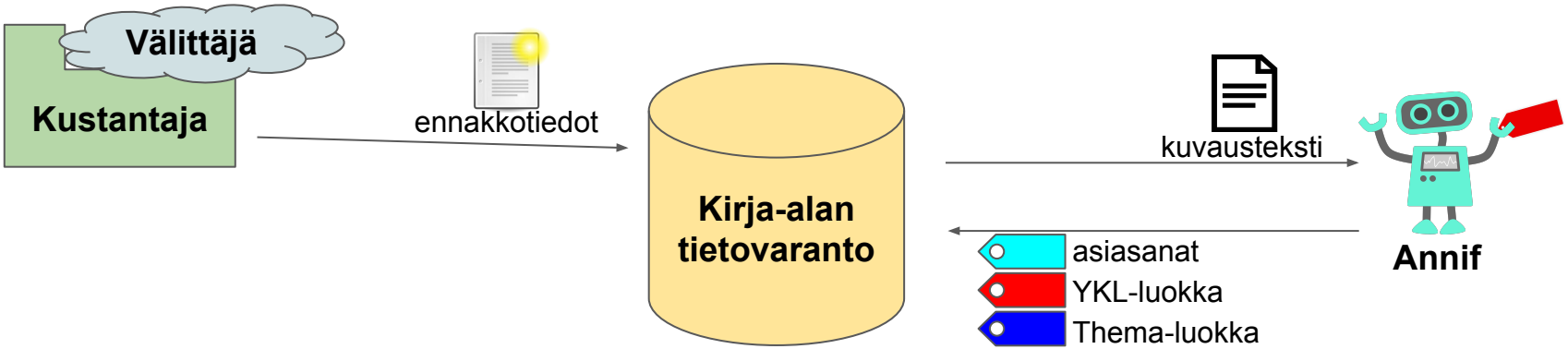
☐ päivitykset takautuvasti

KirjaVälitys

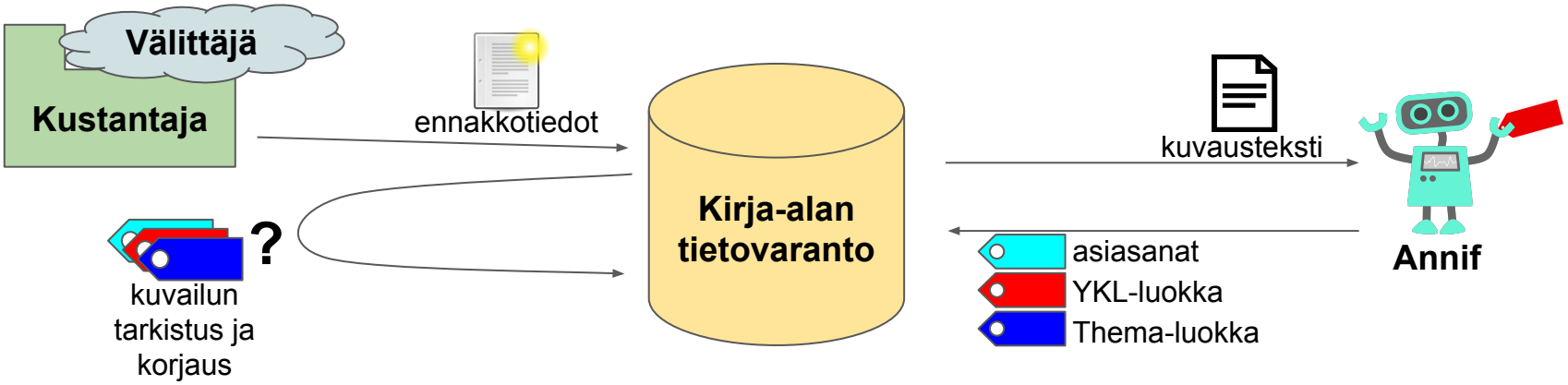
Ennakkotiedon käsittely



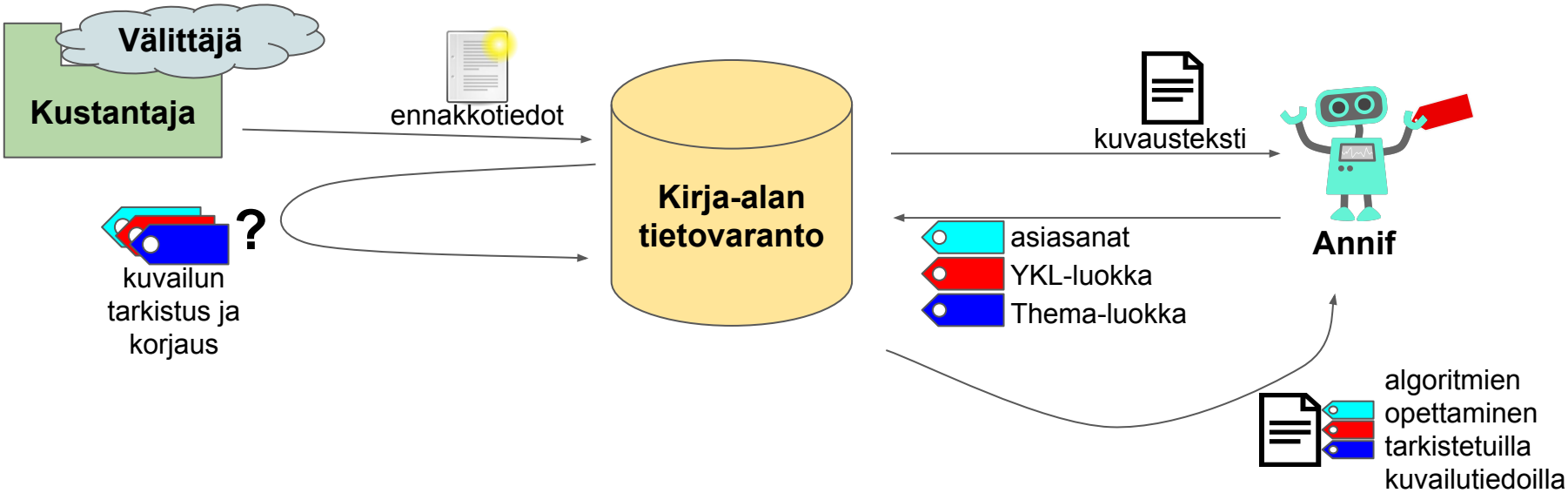
Ennakkotiedon käsittely



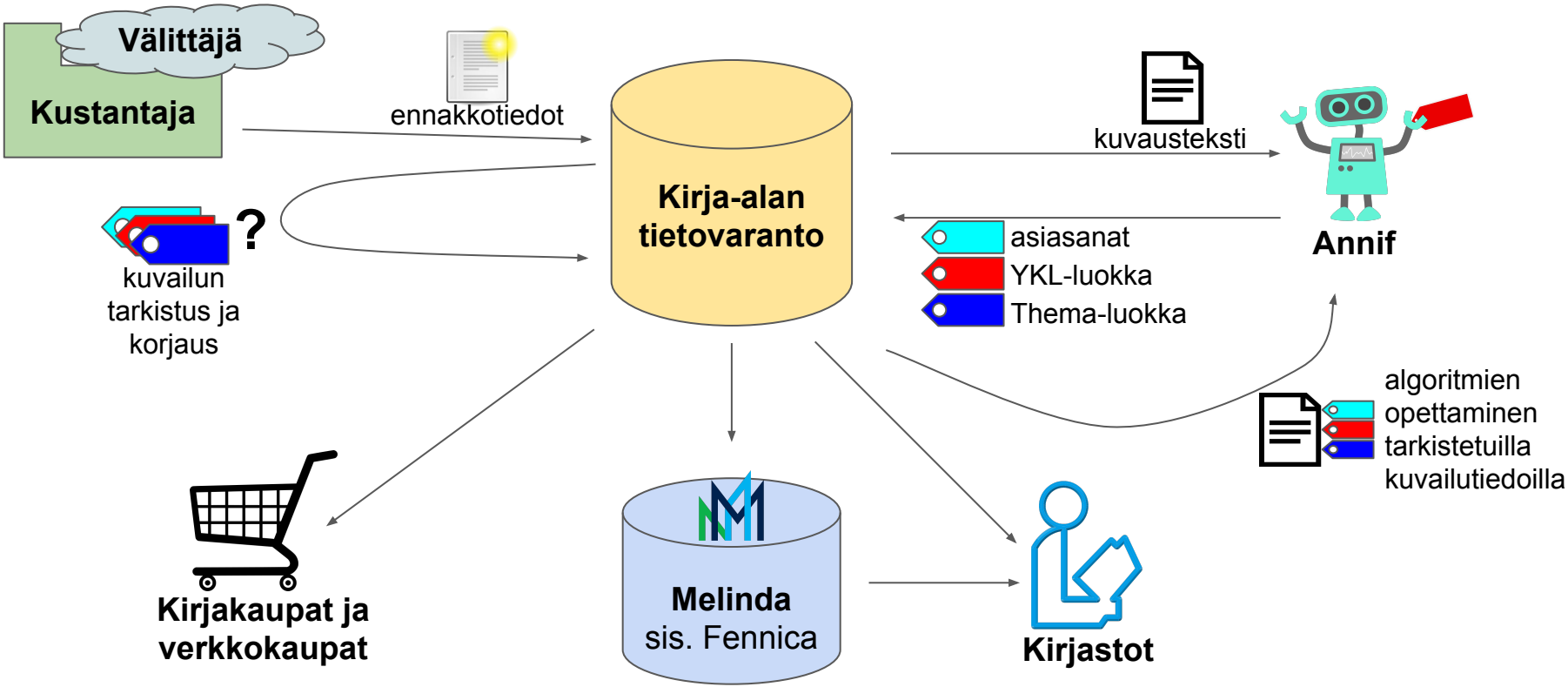
Ennakkotiedon käsittely



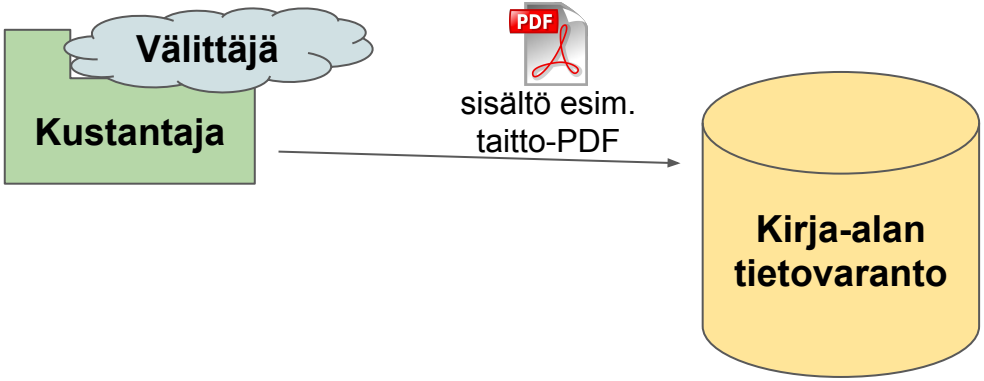
Ennakkotiedon käsittely



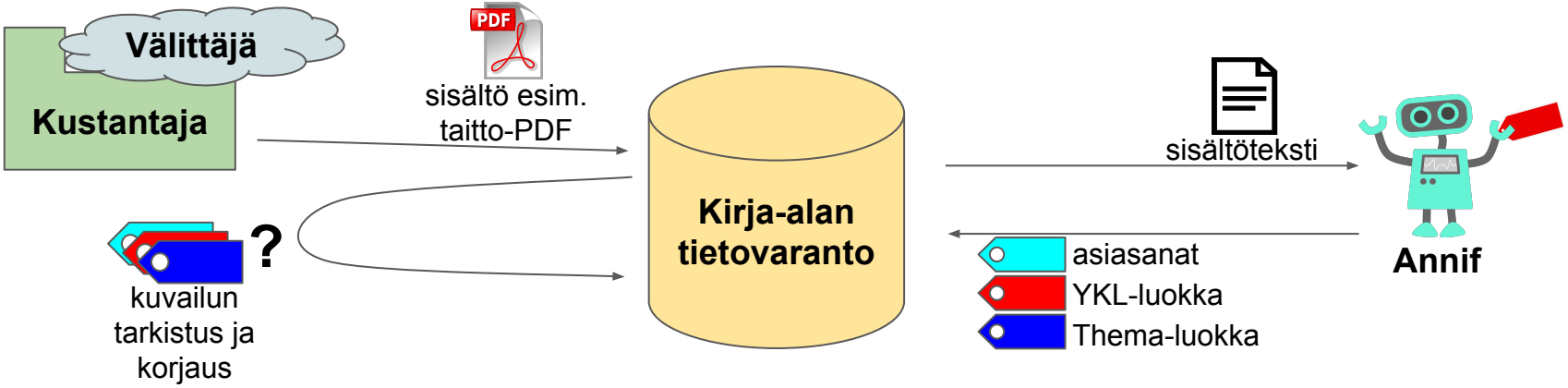
Ennakkotiedon käsittely



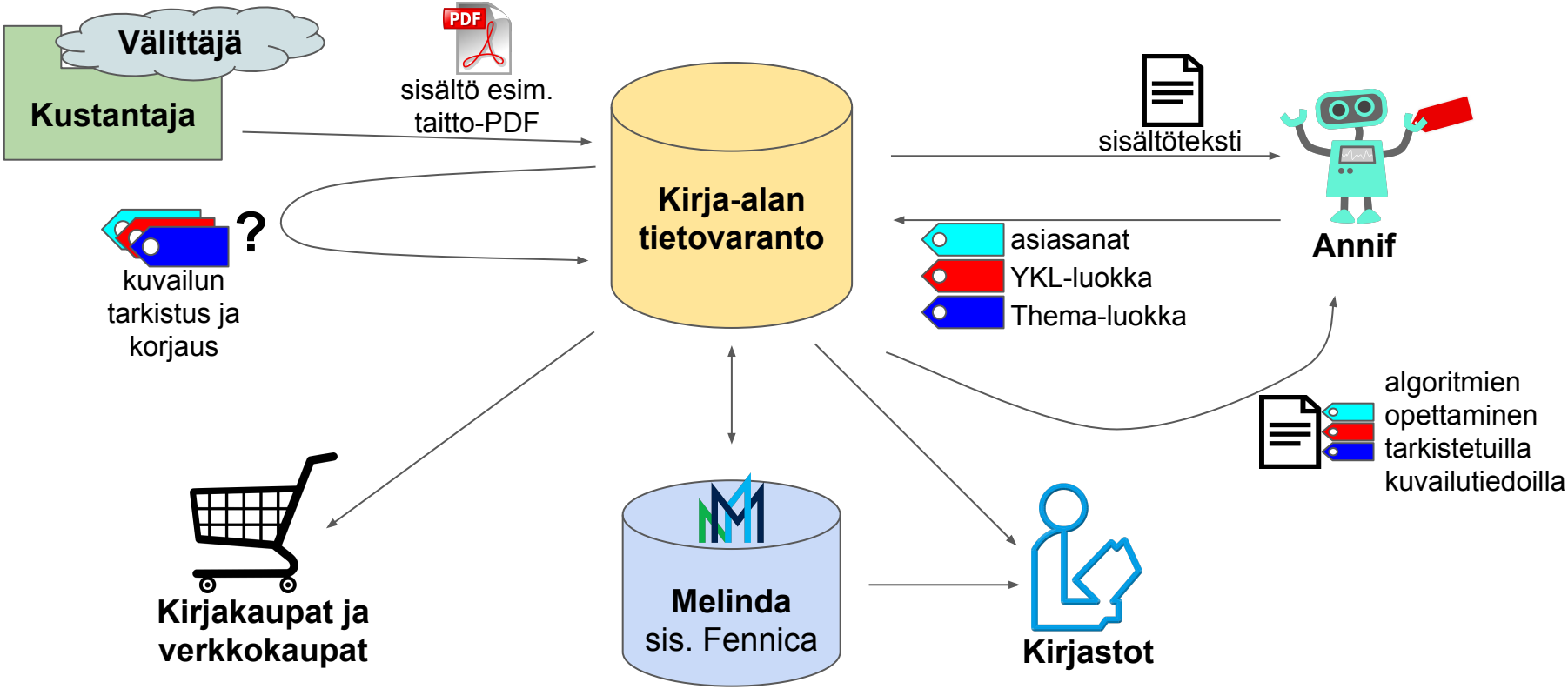
Sisältötarkistus



Sisältötarkistus



Sisältötarkistus



Hyödyt kirja-alan toimijoille

Kustantajat

Mahdollisuus vaikuttaa kuvailuun
- *kaupallisen ja kirjastokuvailun tavoitteiden välillä voi olla ristiriitoja!*

Rikastettu tuotetieto markkinoinnin ja kauppojen käytettävissä

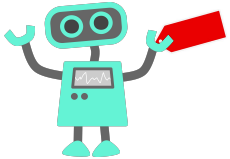
Kirjakaupat ja verkkokaupat

Luotettava tuotetieto (mm. luokitukset) käytettävissä aiemmin

Kirjastot

Luotettava kuvailutieto käytettävissä aiemmin

Mahdollista tehdä kuvailu aiempaa automaattisemmin ja aiemmin
(ei tarvitse odottaa vapaakappaleita?)



Annifin käyttäjät

Algoritmit oppivat palautteen kautta tuottamaan parempaa asiasanoitusta ja luokitusta kaikkien käyttöön

Kiitos!

Kysymyksiä?

osma.suominen@helsinki.fi - @OsmaSuominen

http://annif.org

Tämä esitys: https://tinyurl.com/annif-metatieto