



**VOIKO TEKNOLOGIA PARANTAA
TERVEYSTIEDON JA -KOMMUNIKAATION
YMMÄRRETTÄVYYTTÄ?**

Hanna Suominen, Tapio Salakoski, Filip Ginter
Turun yliopisto

IKITIK-teemapäivä 16.9.2009, Turku

1. Johdanto

1. Informaatio- ja kieliteknologia
2. Terveystieto ja -kommunikaatio

2. Ymmärrettävyys ja tavoitettavuus

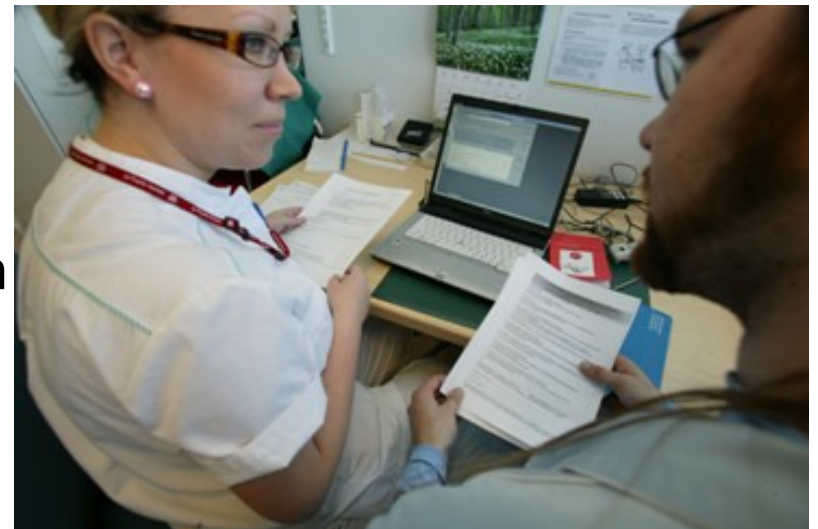
1. Ongelmat nykytilassa
2. Tukemisella saavutettavat edut

3. Teknologisia komponentteja

1. Tulkitseva tiedonhaku
2. Diagnosointi
3. Suomen kielen ymmärtäminen

4. Komponenteista kokonaisuus

1. Tiedonkulun kokonaisuus
2. Iteratiivinen prosessi
3. Toimijoiden yhteistyö



Kuva: Turun yliopiston viestintä



INFORMAATIO- JA KIELITEKNOLOGIA

Menetelmiä ja ratkaisuja puhutun tai kirjoitetun kielen tunnistamiseen, tuottamiseen ja analysoimiseen

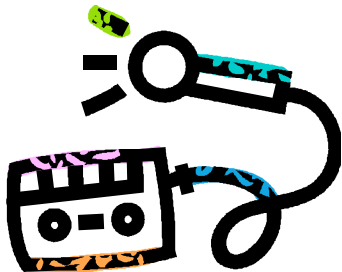
Esimerkiksi

- Puheentunnistus
- Sanastollinen ja kieliopillinen tarkistus ja yhdenmukaistaminen
- Sisällön ohjaaminen rakenteellisesti ja alueellisesti
- Ryhmittely ja strukturoidun tiedon automaattinen tuottaminen
- Tulkitseva tiedonhaku
- Lomakkeiden täyttäminen
- Tiivistäminen

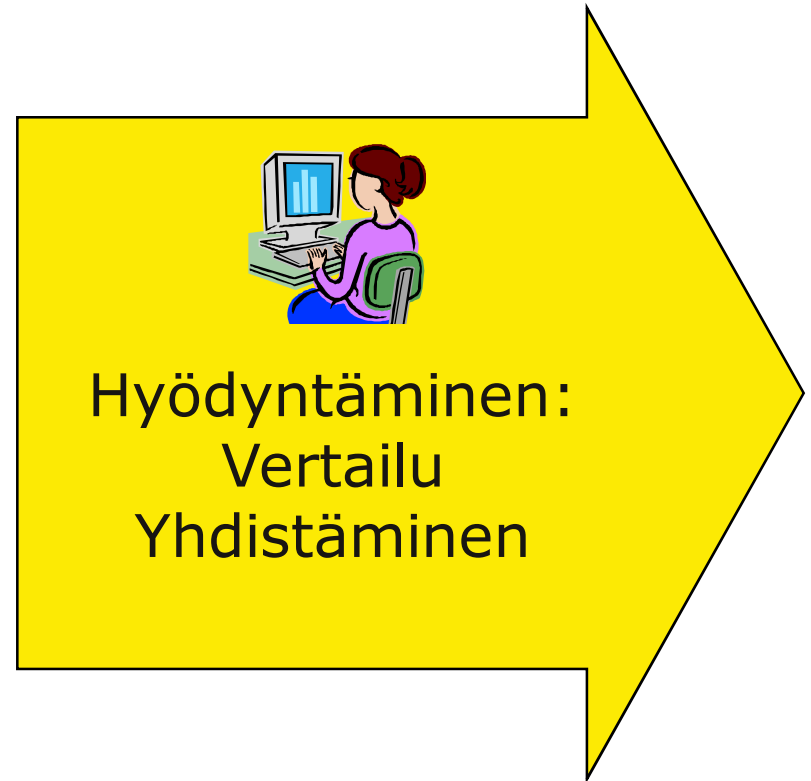
- Numeerinen ja luokittainen



- Puhuttu ja kirjoitettu teksti



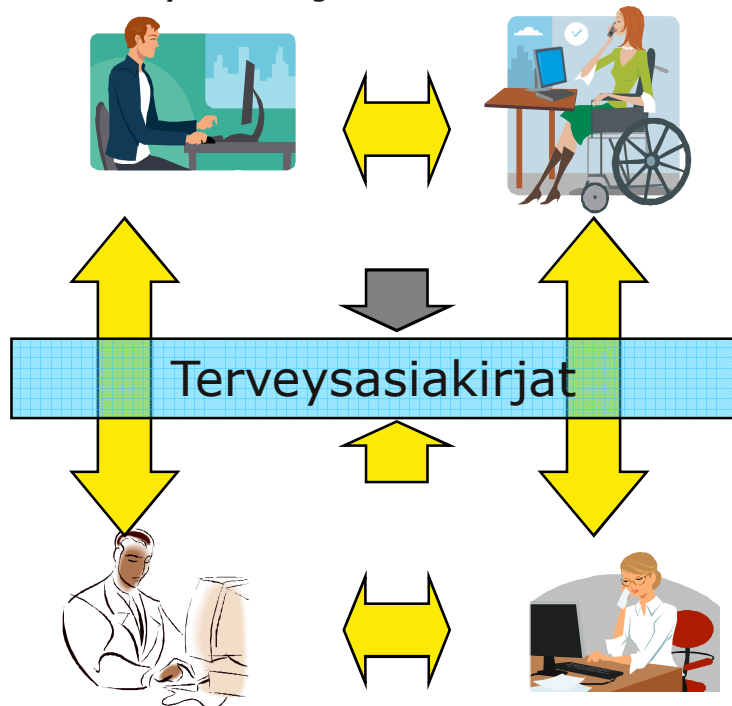
*Pitkä aamuv
Teholle tulua
nopea fa, jota
yritetty kääntää
sähkllä (x3)
tuloksetta.
Diureesi heikko,
aamulla ...*



- Monikielinen ja -ammattillinen ympäristö, jossa tieto- ja kommunikaatiotarpeet vaihtelevat

Henkilöasiakkaat:

Terveyden- ja elämänhallinta



Kommunikaation kielet

Suomi, ruotsi, englanti, ...

Turku, savo, uusimaa, ...

Teho-osasto, sisätaudit, työterveys, ...

Sairaanhoitaja, lääkäri, radiologi, ...

Moniammatillinen terveystalvelujen toimittajatiimi:
Kliininen päätöksenteko ja hallinto

- Terveystiedon ongelmat
 - Suuri määrä
 - Hajanainen rakenne
 - Referentiaalisen moninaisuus
- Ymmärrettävyyden ja tavoitettavuuden edut
 - Aktivoivat henkilöasiakkaita edistämään terveyttä entistäkin paremmin
 - Omakielisyys ja yhteydet lisätietoon
 - Nopeuttavat terveyskommunikaatiota
 - Tietotarpeet ja välittyvyys
 - Ennaltaehkäisevät hoitovirheitä
 - Tiedon saatavuus ja oikeellisuus

- Lehtikunnas, Salanterä & Hupli (2002). Kirjaaminen tehohoitotyössä. Yksilöiden terveyden ja hyvinvoinnin vahvistaminen eri ympäristöissä elämänkulun kaikissa vaiheissa VII kansallinen hoitotieteen konferenssi
- Suominen, Lehtikunnas, Back, Karsten, Salakoski, Salanterä (2007). Applying Language Technology to Nursing Documents: Pros and Cons with a Focus on Ethics. Int J Med Inform 76S2
- Suominen, Lundgrén-Laine, Perttilä, Salakoski, Salanterä (2008). Tehohoidon elektroniset potilasasiakirjat – hyödyntämätön voimavara. Valtakunnalliset Lääkäripäivät 2008

- 516 tehohoitopotilaan asiakirjat

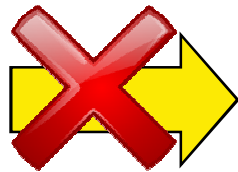
Diagnoosi
Hoitoisuuspiste
Tulotyyppi, LOS

		<i>n</i>	Keski-arvo	Minimi		Maksimi
Tulo	Päiviä	348	250	8	7 s.	2100
Hoito	1	516	270	37		720
	2	516	480	130		1290
	3	516	690	240		1700
	4	516	880	290		2000
	7	382	1500	660		2800
	14	99	2800	1600	48 s.	4600
	Kaikki	516	2100	480		13000
	Keskiarvo	516	190	89		380
Siirto		514	400	140	4 s.	1000

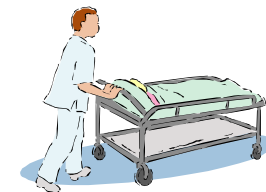
- 516 tehohoitopotilaan asiakirjat
- Monta muotoa samasta otsikosta
 - *Hemodynamiikka*: yli 30 erilaista versiota
- Ovatko otsikot rinnakkaiset?
 - *Hengitys - Hapetus, Hemodynamiikka - Verenkierto*
- Sama tieto eri otsikoiden alla
 - *Hemodynamiikka: ICP - Tajunta: ICP*
- Tiedonkulussa ongelmia



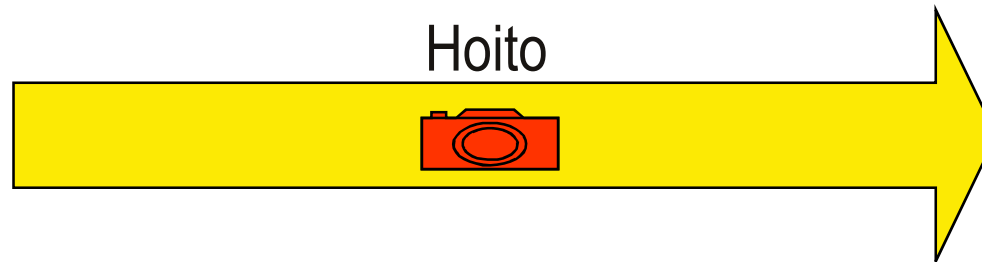
Tulo



Hoito



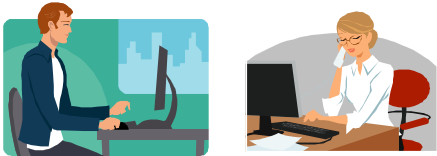
Siirto



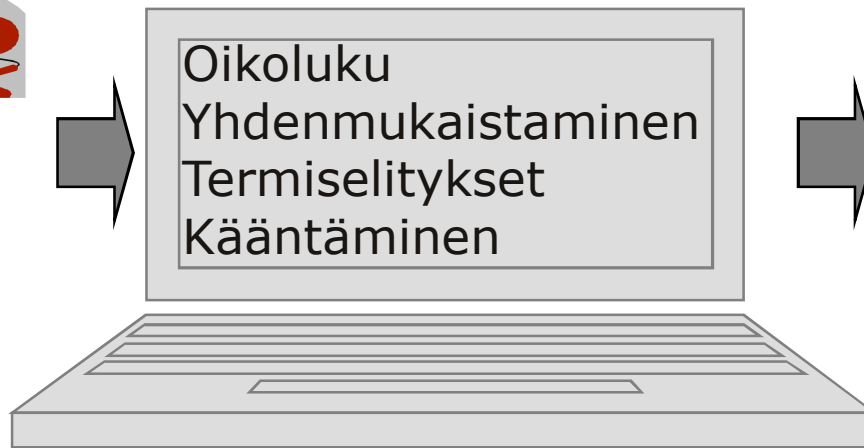
- 516 tehohoitopotilaan asiakirjat
- *Noradrenaliini*
 - Perusmuodolla yli 60 erilaista kirjoitusasua
 - Erilaisia taivutusmuotoja ja johdannaisia yli 350

nonadrenalin, nonadrenalina, nor, norad, noradenaliini, noradenalina, noraderenalina, noraderenalini, noradernalina, nor-adr, nor.adr, noradr, noradranalian, noradranlina, noradreanliini, noradreanlina, noradreliini, noradrelin, noradren, noradrenal, noradrenalaina, noradrenali, noradrenalia, noradrenalian, noradrenaliia, noradrenaliina, noradrenaliimi, noradrenaliin, noradrenaliina, noradrenaliini, noradrenaliinna, noradrenalin, noradrenaline, noradrenalini, noradrenallina, noradrenallini, noradrenanil, noradrenanilin, noradrenenalina, noradrenliini, noradrenlin, noradrenlina, noradrenlini, noradrnaliini, noradrnalina, noradrrenalin, noradrtenalin, norandrenalina, noraradrenalina, norarenaliini, norarenalina, noratrenaliini, nordarenaliini, nordrenaliini, nordrenalin, nordrenalina, nordrenanili, noreadr, norradenalina, norradrenaliini, norradrenalin, norradrenalini, norri, norrradrenalin

- Hoitokertomus englanninkieliselle potilaalle suomalaiselta teho-osastolta

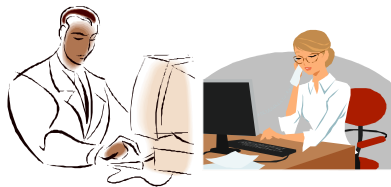


Teholle tulua nopea fa, jota yritetty kääntää sähköllä (x3) tuloksetta.



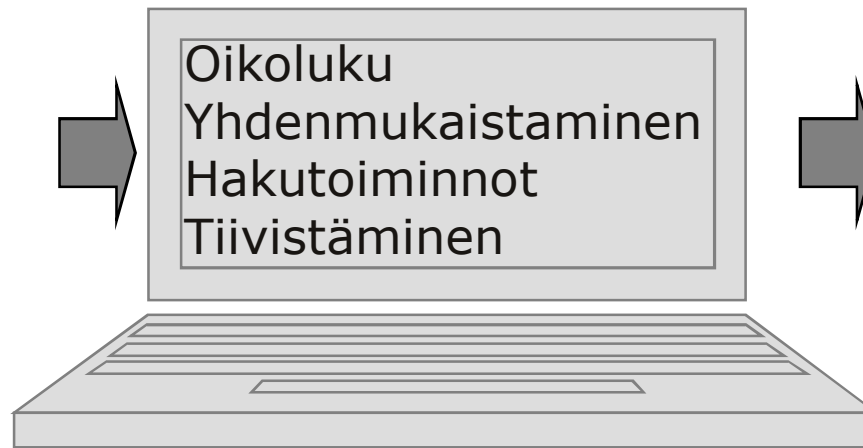
After the intensive care admission, the patient had fast fibrillatio atriorum, which clinicians tried to convert using electricity. Regardless of three attempts, this did not solve the problem.

- Kardiologi tarvitsee tiivistelmän potilaan hemodynamiikasta tehohoitojakson ajalta



Teholle tulua nopea fa, jota yritetty kääntää sähköllä (x3) tuloksetta.

...



Teholle tultua nopeahko FA, jota yritetty kääntää sähköllä (x3) tuloksetta.

Diureesi heikko, aamu LÄÄKE.

...

Poika ja vaimo kylässä.

Aamulla pulssi vaihteleva ja aivopaineet koholla.

Dreenivuoto niukkaa.

- Saatavuus
 - Hakutoiminnot ja tiivistelmien teon tukeminen
 - Strukturoinnin automatisointi
 - Otsikointi
 - Tietokenttien täyttö
 - Tilastointiin tarvitaan diagnoosikoodit, mutta osasta asiakirjoista tieto puuttuu
 - Verrokkimielipide diagnoosista päätöksenteon tueksi tai muistutukseksi
- Oikeellisuus
 - *MRSA v*
 - Kysytty aiheesta? Testi tehty? Tulos positiivinen?
 - Muistutukset kirjata
 - Kivunarviointi ja -hoito

TAVOITE 1: TULKITSEVA TIEDONHAKU

HAKUAIHE: hengitys

ARVIOINTIASTEIKKO:



*omaiset: tytär soitti (kaksi
tytärtä)*

Tajunta: säpsähtelee

HENGITYS yskii silloin tällöin

*kelt. limaa tulee. Reagoi imuihin:
puree intub putkea + yskii.*

*Hapenottokyky ei ok --> o2 2l
viikset -> sopeutuu hyvin*

*Hemodyn. p. /[^] /v, myös hap. /[^]
/v parin min välein. ekg aamulla.*

- Suominen, Pahikkala, Hiissa, Lehtikunnas, Back, Karsten, Salanterä, Salakoski (2006). Relevance Ranking of Intensive Care Nursing Narratives. Lecture Notes in Computer Science 4251
- Hiissa, Pahikkala, Suominen, Lehtikunnas, Back, Karsten, Salanterä, Salakoski (2007). Towards Automated Classification of Intensive Care Nursing Narratives. Int J Med Inform 76S3
- Suominen, Lundgrén-Laine, Salanterä, Salakoski (2009). Evaluating Pain in Intensive Care. Studies in Health Technology and Informatics 146
- Ginter, Suominen, Pyysalo, Salakoski (2009). Combining Hidden Markov Models and Latent Semantic Analysis for Topic Segmentation and Labeling: Method and Clinical Application. Int J Med Inform. In Print

- Aiheet: hengitys, verenkierto, kipu
- 3 asiantuntijaa
- Pilottiaineisto suomalaisilta teho-osastoilta

Cohenin κ (95% CI)	Hengitys	Verenkierto	Kipu
$N_1 - N_2$,73 (,68–,78)	,89 (,85–,92)	,88 (,82–,94)
$N_1 - N_3$,67 (,62–,72)	,81 (,77–,86)	,79 (,73–,86)
$N_2 - N_3$,82 (,82–,89)	,87 (,83–,90)	,76 (,69–,83)

Matemaattinen optimointiongelma

$$\min_f \sum_i (f(x_i) - y_i)^2 + \lambda \|f\|_k^2, \text{ missä}$$

$$k(x, x_i) = \frac{\langle x, x_i \rangle}{\sqrt{\langle x, x \rangle \langle x_i, x_i \rangle}}$$

Lingvistinen käsittely: Snowball stemmer suomelle

*Teholle tultua nopeahko →
teho tul nopeahko*



ikitik AUTOMAATTINEN LUOKITTELU

AUC (95% LV)	Hengitys			Verenkierto			Kipu		
$C(N_1)$ vs. N_1	,86 (,82–,90)			,89 (,84–,93)			,71 (,61–,80)		
$C(N_2)$ vs. N_2	,88 (,85–,91)			,93 (,90–,97)			,81 (,73–,89)		
$C(N_3)$ vs. N_3	,87 (,84–,91)			,91 (,86–,95)			,71 (,61–,80)		

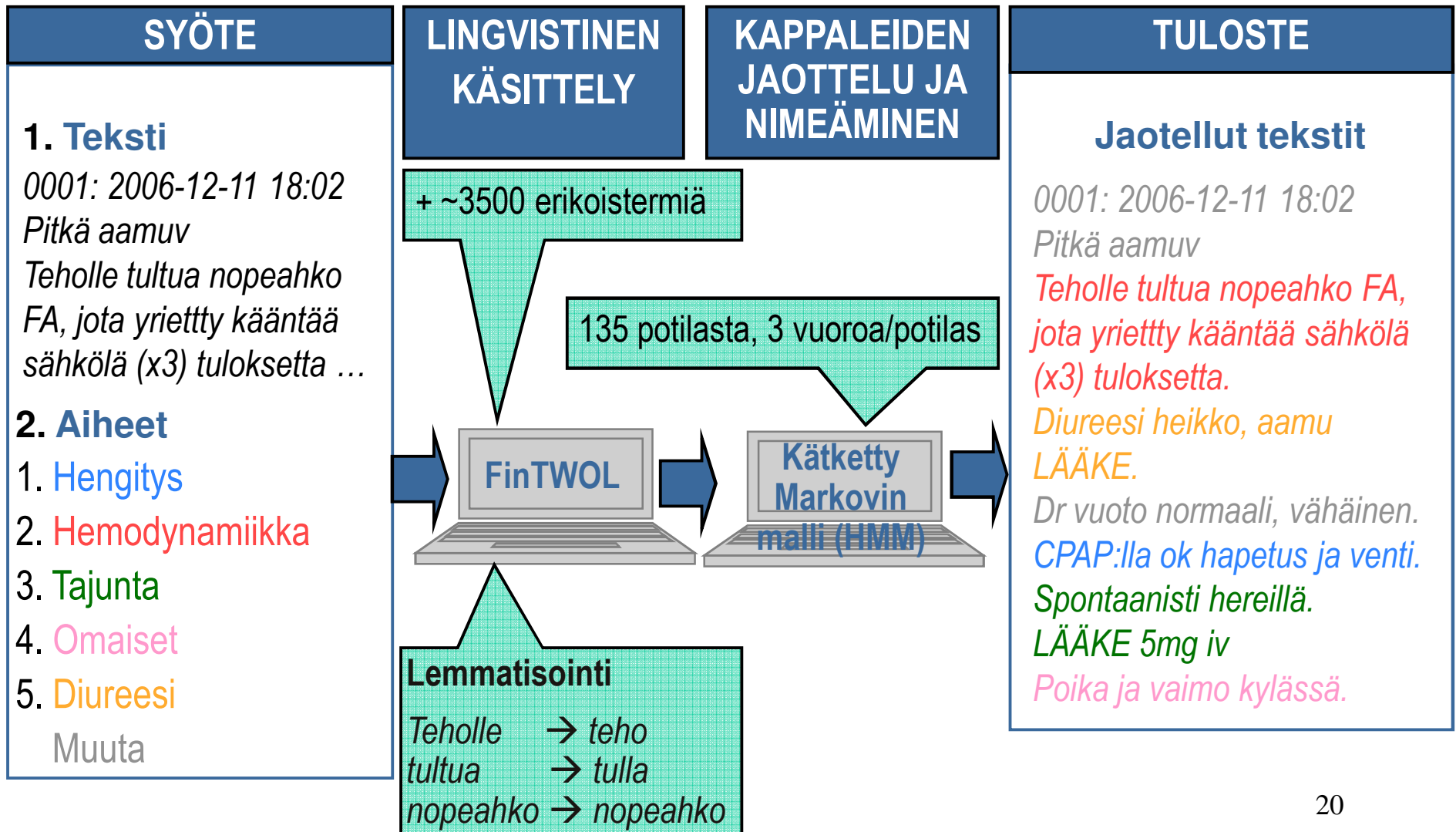
AUC	Hengitys			Verenkierto			Kipu		
	N_1	N_2	N_3	N_1	N_2	N_3	N_1	N_2	N_3
$C(N_1)$,86	,74	,72	,89	,93	,91	,71	,81	,72
$C(N_2)$,83	,88	,86	,88	,93	,91	,71	,81	,71
$C(N_3)$,84	,88	,87	,89	,93	,91	,67	,77	,71

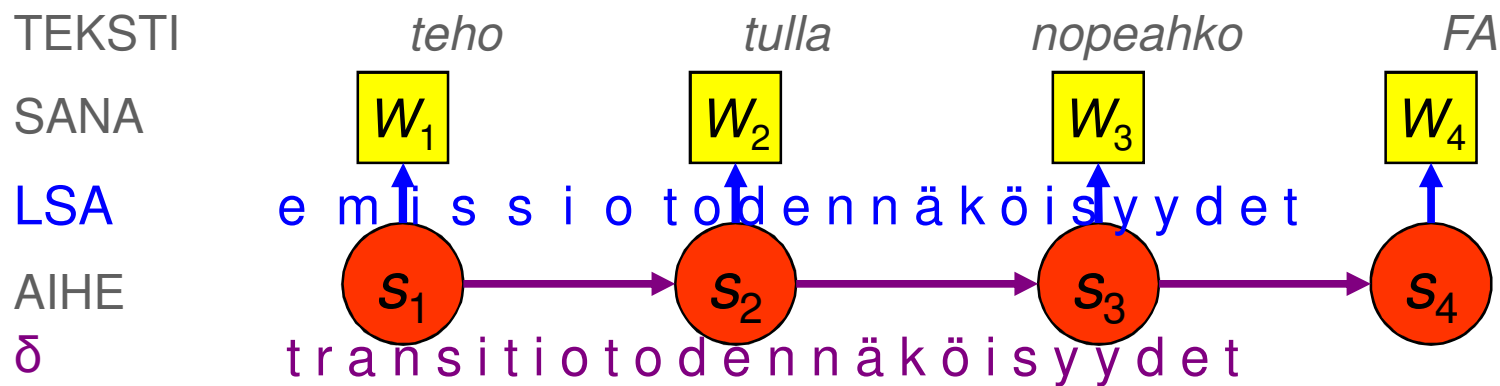
RELEVANSSIN ARVIOINTI

	3	2	1	0
Hengitys	Hapetus Hengityskoneet	Lima Yskä Pleuradreenit		
Verenkierto	Pulssi Verenpaine	Lämpö Ihonväri		
Kipu	<i>kipu</i> <i>särky</i>	Epäsuorat kipuindikaattorit		Ei <i>kipu</i>

	Kendallin τ_b	95% LV
Hengitys	,62	(,56–,68)
Verenkierto	,69	(,61–,76)
Kipu	,44	(,30–,59)

KAPPALEIDEN AIHEET JA KAPPALEENJAOT AUTOMAATTISESTI





Schützen Word-space LSA:

- Sanan ja aiheen samankaltaisuus
- $P(\text{sisko}, \text{omaiset}) \gg P(\text{sisko}, \text{hengitys})$

Jatketaan samaa aihetta

- Suuri $\delta \rightarrow$ suuret kappaleet: *hemodynamiikka*
- Pieni $\delta \rightarrow$ pienet kappaleet: *diureesi*

Word sequence: $w = (w(1), \dots, w(T))$ Words: $w_j, j \in \{1, \dots, N_w\}$

Topic sequence: $q = (q(1), \dots, q(T))$ Topics: $q_i, i \in \{1, \dots, N_q\}$

other = other NOT breathing NOT hemodynamics NOT consciousness NOT relative NOT diuresis

$q_6 = q_6$ NOT q_1 NOT q_2 NOT q_3 NOT q_4 NOT q_5

Problem: $\hat{q} = \arg \max_q P(w(1) | q(1)) \prod_{t=2}^T P(w(t) | q(t)) P(q(t) | q(t-1))$

Emission probabilities: $P(w_j | q_i) = \text{LSA}(w_j, q_i)$

Transition probabilities: $P(q_j | q_i) = \begin{cases} \delta, & \text{if } j = i \\ \frac{1 - \delta}{N_q - 1}, & \text{if } j \neq i \end{cases}$

➔ $\hat{q} = \arg \max_q \text{LSA}(w(1), q(1)) \prod_{t=2}^T \text{LSA}(w(t), q(t)) P(q(t) | q(t-1))$

$$\hat{q} = \arg \max_q \text{LSA}(w(1), q(1)) \prod_{t=2}^T \text{LSA}(w(t), q(t)) P(q(t) | q(t-1))$$

Arvojen vertailun ongelma

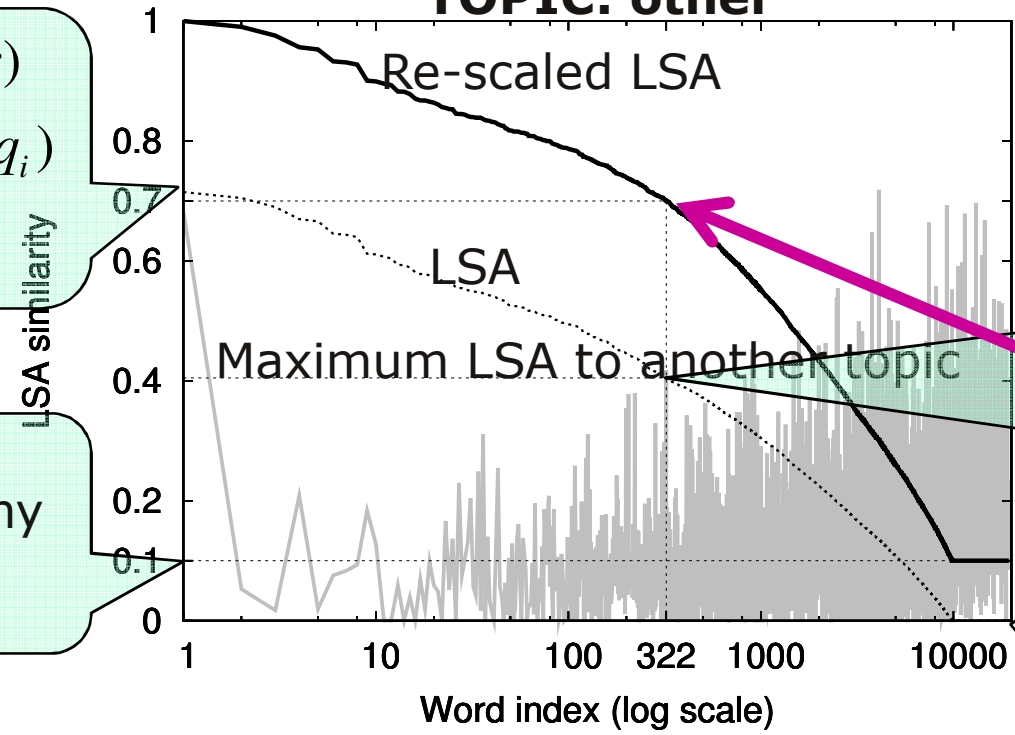
OMAISET	HEMODYNAMIIKKA	MUUTA	
<i>omainen</i> 1.000	<i>hemodynamiikka</i> 1.000	<i>vatsa</i>	0.683
<i>soittaa</i> 0.947	<i>pulssi</i> 0.910	<i>muu</i>	0.682
<i>tytär</i> 0.916	<i>sr</i> 0.819	<i>verkko</i>	0.676
<i>vaimo</i> 0.889	<i>rr-taso</i> 0.785	<i>hemolysoituu</i>	0.673
<i>vierailu</i> 0.877	<i>korkeahko</i> 0.784	<i>paita</i>	0.637
<i>poika</i> 0.859	<i>sinusrytmi</i> 0.784	<i>varjoainetehoste</i>	0.635
<i>kello</i> 0.821	<i>rr</i> 0.768	<i>peite</i>	0.630
<i>mies</i> 0.820	<i>verenpaine</i> 0.716	<i>LÄÄKKEEstä</i>	0.618
<i>veli</i> 0.785	<i>lisälyönti</i> 0.673	<i>pehmeä</i>	0.618
<i>sisko</i> 0.777	<i>ok</i> 0.672	<i>neulanäyte</i>	0.614

LSA-ARVOJEN SKAALAUUS

TOPIC: other

$LSA_1(q_i = other)$
 $= \max_{w_j} LSA(w_j, q_i)$
 $= 0.71$

Minimal
 similarity of any
 word to any
 topic β



$I(q_i = other) = 322$
 $LSA_I(q_i = other)$
 $= 0.41$
 $-- > \alpha = 0.7$
 (topic-independent)

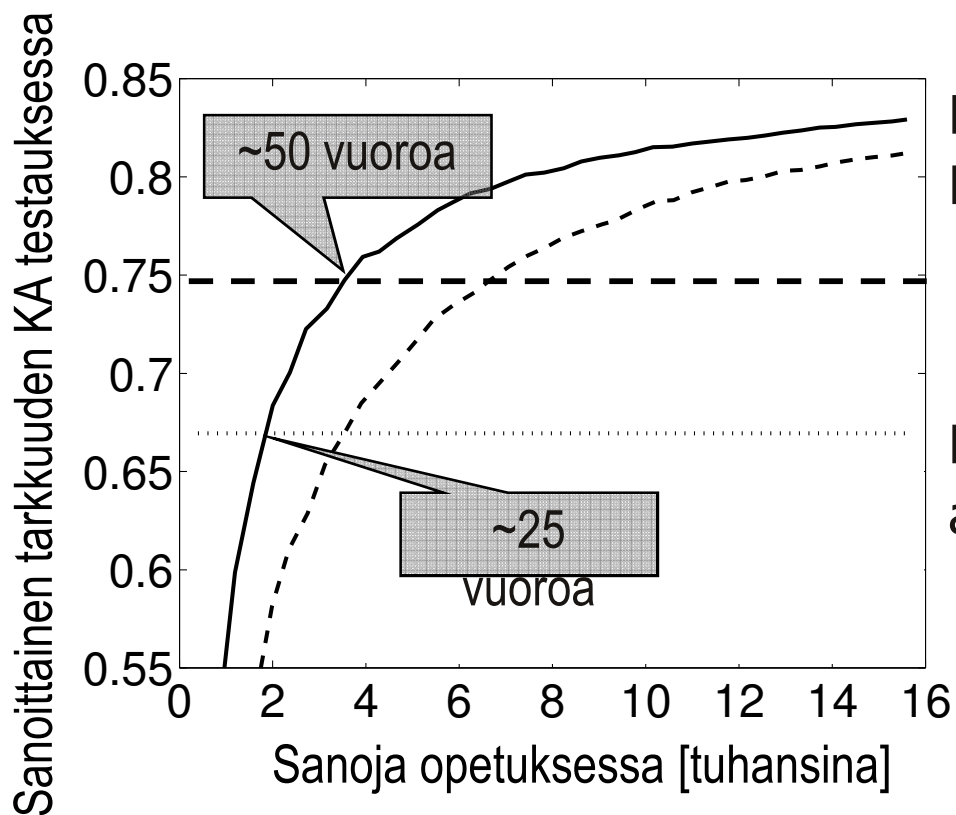
$LSA_m(q_i = other)$
 $= \min_{w_j} LSA(w_j, q_i)$
 $= 0$

$$\hat{q} = \arg \max_q \text{LSA}'(w(1), q(1)) \prod_{t=2}^T \text{LSA}'(w(t), q(t)) P(q(t) | q(t-1)),$$

where

$$\text{LSA}'(w_j, q_i) = \begin{cases} \frac{(1-\alpha)(\text{LSA}(w_j, q_i) - \text{LSA}_I(q_i))}{\text{LSA}_I(q_i) - \text{LSA}_I(q_i)} + \alpha, & \text{if } w_j \text{ is before } I(q_i) \\ \frac{(\alpha-\beta)(\text{LSA}(w_j, q_i) - \text{LSA}_m(q_i))}{\text{LSA}_I(q_i) - \text{LSA}_m(q_i)} + \beta, & \text{if } w_j \text{ is before } I(q_i) \end{cases}$$

HMM JA LSA-HMM MENETELMIEN VERTAILU



Lingvistinen käsittely & HMM

HMM

Lingvistinen käsittely & LSA-HMM

Lingvistinen käsittely +

Erilaiset tietotarpeet



Menetelmän valinta



TULOKSIA 1: TULKITSEVA TIEDONHAKU

- Tavoite
 - Tekstistä voidaan hakea tietoa aiheen perusteella ymmärtäen ja tulkiten
 - Käyttäjä määrittää ja ohjaa hakua
- Tulosten käyttökelpoisuus
 - Laatu ja luotettavuus
 - Nopeus ja implementoitavuus



TAVOITE 2: AUTOMAATTINEN DIAGNOSOINTI

CLINICAL HISTORY

This is a 7-month - old male with wheezing.

IMPRESSION

Borderline hyperinflation with left lower lobe atelectasis versus pneumonia. Clinical correlation would be helpful. Unless there is clinical information supporting pneumonia such as fever and cough, I favor atelectasis.

ICD-9-CM CODING

486	Pneumonia, organism unspecified
518.0	Pulmonary collapse
786.07	Wheezing

45 koodia
94 yhdistelmänä



TUTKIMUKSIA

- Suominen, Ginter, Pyysalo, Airola, Pahikkala, Salanterä, Salakoski (2008). Machine Learning to Automate the Assignment of Diagnosis Codes to Freetext Radiology Reports: a Method Description. Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health-Care Applications
- Farkas, Szarvas (2008). Automatic construction of rule-based ICD-9-CM coding systems. BMC Bioinformatics 9S3

PIIRTEIDEN JALOSTAMINEN

45 keinokekoista esimerkkiä
Tokenisointi
UMLS-käsitteet ja -yläkäsitteet
Konteksti
Binääripiirteet



LUOKITTELU

RLS

virhe

Ripper

45 koodia
94 yhdistelmänä

Keinotekoiset esimerkit opetuksessa

580-629 Diseases of the genitourinary system

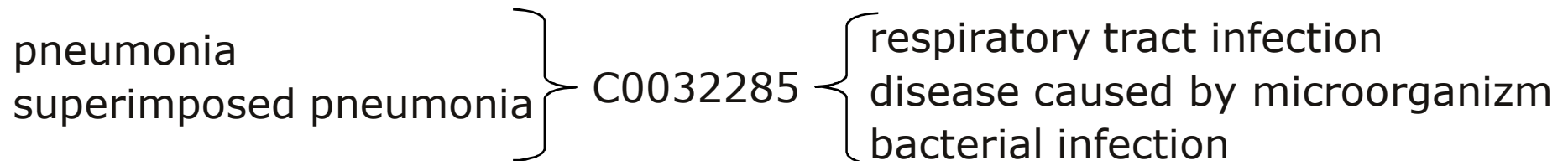
590-599 Other diseases of urinary system

593 Other disorders of kidney and ureter

593.7 Vesicoureteral reflux

593.70 Vesicoureteral reflux unspecified or without reflux nephropathy

UMLS-käsitteet ja -yläkäsitteet



Konteksti: negaatiot ja spekulatiiviset ilmaukset

*no, positive, suggestive, likely**no pneumonia**no respiratory tract infection*

3 asiantuntijaa

Asiantuntija - Asiantuntija

Asiantuntija - Gold standard

F1-mikro

,673–,758

,826–,896

**Noin 150 osallistujaa,
44 valmista järjestelmää**
6 maanosaa ja yli 20 maata

	F1-mikro
Keskiarvo	,766
Keskihajonta	,133
Mediaani	,799
1.	,891
2.	,886
3.	,877
4.	,876
5.	,872

Järjestelmämme opetusaineistolla**Komponentti****F1-mikro**

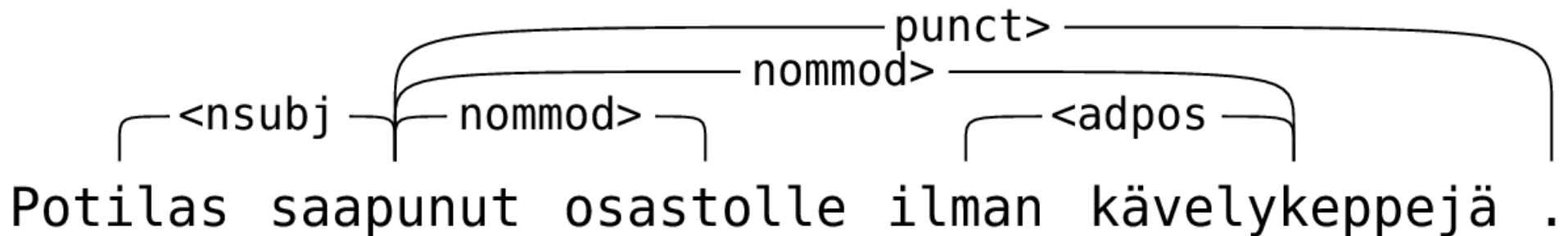
RLS	,793
Tokenisointi	,807
UMLS-käsitteet	,825
UMLS-yläkäsitteet	,834
Konteksti	,847
Jatkokäsittelyyn RIPPER	,865
Keinotekoiset esimerkit	,866



TULOKSIA 2: AUTOMAATTINEN DIAGNOSOINTI

- Tavoite
 - Diagnoosikoodien muodostaminen tekstisisällön perusteella
 - Tilanteisiin, joissa koodeja ei ole annettu tai päätöksenteon tueksi
- Tulosten käyttökelpoisuus
 - Laatu ja luotettavuus
 - Nopeus ja implementoitavuus
 - Kielispesifiset resurssit

- Jäsentämisellä selville sanan rooli suhteessa muihin lauseen sanoihin
- Ymmärrys lauseesta kasvaa. Se ei ole enää vain sanajono.
- Jäsentämisellä on keskeinen rooli lauseen merkityksen ymmärtämisessä



- **Kuka** saapui? Potilas (subjekti)
- **Minne?** Osastolle (nominaalimääre allatiivissa)
- **Kenellä** ei ollut kävelykeppejä? (potilas vai osasto?)

- Haverinen, Ginter, Laippala, Salakoski (2009). Parsing Clinical Finnish: Experiments with Rule-Based and Statistical Dependency Parsers. Proceedings of NODALIDA'09.
- Laippala, Ginter, Pyysalo, Salakoski (2009). Towards Automated Processing of Clinical Finnish: A Sublanguage Analysis and a Rule-Based Parser. Int J Med Inform. In Print

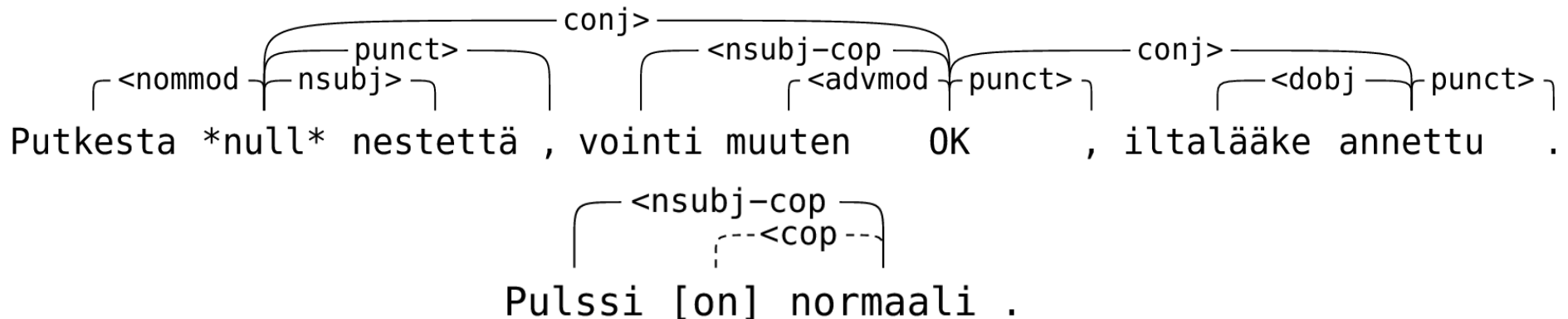
- Miksi
 - Sanajärjestyksen normalisointi
 - *Potilas saapui osastolle. Osastolle saapui potilas.*
 - Ymmärtäminen
 - Mitä on tehty (yleensä pääverbi)
 - Ketä/Mitä (yleensä subjekti)
 - Kenelle/Mitä (yleensä objekti)
- Tärkeä analysointivaihe
 - Ei jäsennetä jäsentämisen takia.
 - Hyödyksi esimerkiksi
 - Tiivistämisessä
 - Kääntämisessä
 - Tulkitsevassa tiedonhaussa



JÄSENTÄMINEN

- Voidaan tehdä automaattisesti
 1. Sääntöpohjaisesti ohjelmoiden kaikki kielioppisäännöt
 2. Tilastollisella koneoppimisella jäsentäen ensin riittävän suuri määrä lauseita (n. 10 000) ja käyttäen niitä sitten "koneen" opetusimerkkeinä
 - Molemmat tavat vaativat paljon työtä – ei ilmaisia lounaita
- Melko hyvä tarkkuus yleiskielellä
 - Monilla kielillä yli 80% riippuvuuksista oikein
 - ... Japani 90%, mutta turkki 64%
 - Saavutettava tarkkuus: kieli ja opetusaineiston koko

- Eräät erityispiirteet on huomioitava
 - Paljon kirjoitusvirheitä
 - Paljon pois jätettyjä lauseenjäseniä
 - Joskus pääverbi on jätetty pois → jäsentimen on lisättävä "null"-sana
- Toisaalta
 - Kielioppi melko rajoitettu ja lauserakenne yleensä helpohko
 - Lauseet paljon lyhyempiä kuin yleiskielessä (KM 7 sanaa, sanomalehtitekstissä KM 10 sanaa ja Wikipediassa KM 14)





KLIINISEN SUOMEN KIELEN JÄSENTÄMINEN

Laippala ym.

- Sääntöpohjainen jäsennin
- Vaikeuksia valita oikea jäsenitys
 - Tuottaa lauseesta noin 20 jäsenitystä, jotka ovat uskottavia syntaktisesti mutta eivät semanttisesti.
- Tällä hetkellä keskitymme jäsentämisessä tilastolliseen koneoppimiseen

Haverinen ym.

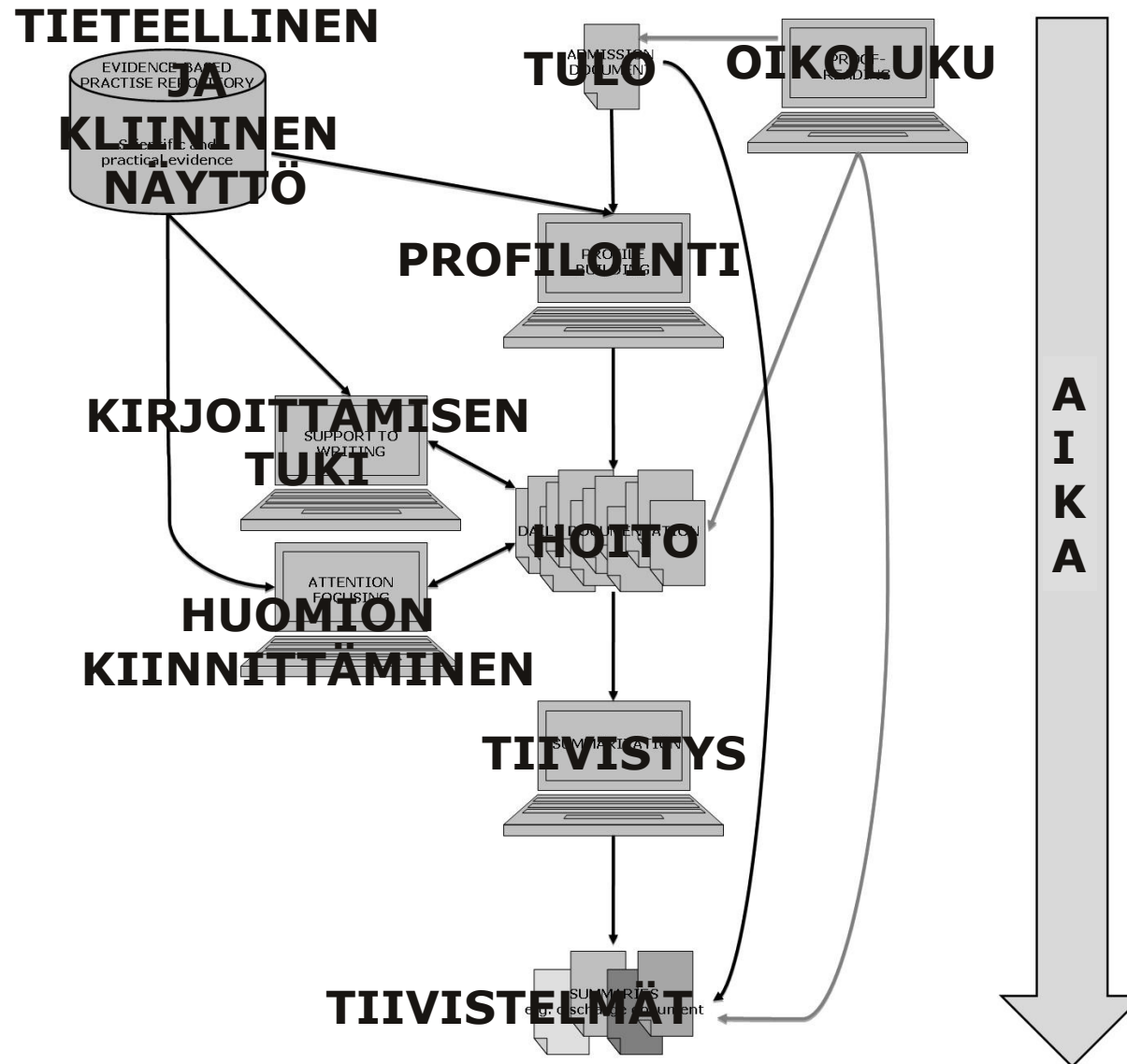
- 1100 annotoitua lausetta
- Tilastolliseen koneoppimiseen perustuva jäsennin
- Tarkkuus noin 70%
- Opetusaineiston pieneen kokoon nähden yllättävän hyvä
- Ei kuitenkaan vielä riittävän hyvä sovelluksiin
- Jatkotutkimus

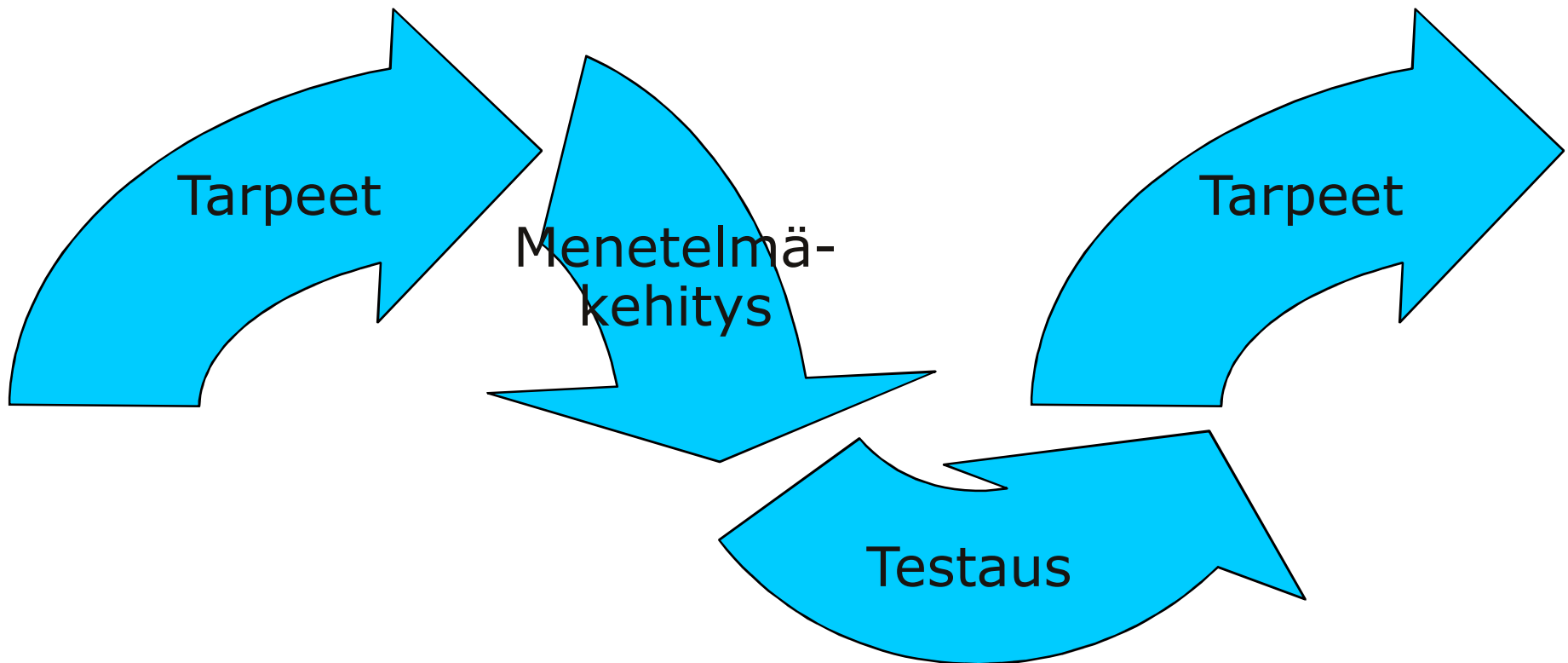


TULOKSIA 3: SUOMEN KIELEN YMMÄRTÄMINEN

- Tavoite
 - Kielellinen ymmärtäminen
 - Kielen rakenne ja terminologiat
 - Kielispesifisten resurssien ja menetelmien tuottaminen
- Lingsoft Proofreader for Health –
suomen kielen tarkistin kliiniselle tekstille
- Tulosten käyttökelpoisuus
 - Laatu ja luotettavuus
 - Nopeus ja implementoitavuus
 - Kielispesifiset resurssit!

KOMPONENTEISTA KOKONAISUUKSIIN





ALKU: Tarvemäärittely ja vaikutusten arviointi aidossa käyttöympäristössä

VERTAILU: Tieteellinen ja kaupallinen huipputeknologia

LOPPUTULOS: Etua yksilölle, yrityksille, yliopistoille ja yhteiskunnalle



YTY

Yksilöllistä Terveyttä yhdessä Ymmärtäen

HEXAnord

HEalth teXt Analysis network
in the Nordic and Baltic countries

