

Digital inkludering
i det uppkopplade samhället
för grupper med speciella behov

Arne Jönsson

Linköpings universitet och RISE SICS East



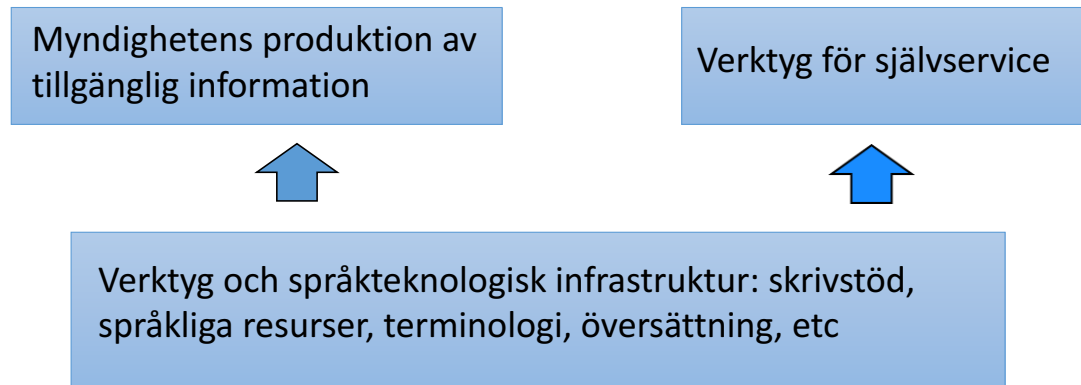
Inkludering av alla medborgare i det digitala samhället

- Utlandsfödda (1.5 milj invånare + nyanlända) har stora problem att använda myndigheternas e-tjänster
- Äldre seniorer och personer med kognitiva funktionshinder klarar inte att navigera på webben.
- Personer med dyslexi eller nedsatt syn undviker svårlästa texter på webben
- 25% av ungdomarna har efter genomgången grundskola svårigheter att förstå en nyhetsartikel

F DINA



Tillgänglig och begriplig myndighetsinformation



F DINA



1. Stöd för myndighetens webbproduktion

- Skribenter (enhetlig o rekommenderad terminologi, läsbarhet, textförenklingar, kvalitetskontroll, skrivregler, etc).
- Underhåll och uppdatering av information, t ex efter ändringar i regler och förordningar.
- Terminologiunderhåll (skörda och klustra termkandidater med möjliga definitioner).
- Flerspråkig myndighetsspecifik terminologi (både auktoriserade och vardagstermer med översättningar).
- Översättningar (egenutvecklade verktyg samt standardverktyg plus kundanpassning med verksamhetsområdets terminologi)

2. Stöd för individualiserad självservice

- Förenklande omskrivning av texter för bättre förståelse
- Navigering och sökning med sammanfattningar
- Översättningar med områdesspecifik flerspråkig terminologi
- Visualisering av text med internationellt symbolspråk

F DINA



Projektdeltagare

- Akademi/forskningsinstitut: RISE SICS East, Linköpings universitet
- Företag: Fodina Language Technology AB, Convertus AB, Briteback AB, Symbolbruket AB.
- Myndigheter: SLL 1177 Vårdguiden/Invånartjänster, Region Östergötland, Linköpings kommun, Linköpings universitet.
- Slutanvändare: Kommunikatörer, utlandsfödda, personer med läs- och förståelsesvårigheter, äldre seniorer, tolkar, . . .



Stödjande plattformar och verktyg

- Convertus – Maskinöversättningsstöd med kundanpassning av terminologi och språkbruk
- Fodina Acrolinx – Skrivstöd, underhåll av dokument och terminologi, mm.
- TeCST – Texttransformationer för bättre läsbarhet och förståelse
- Symbolbruket – Internationellt symbolspråk för annotering av text
- Briteback – Kommunikationsplattform för självservicetjänster

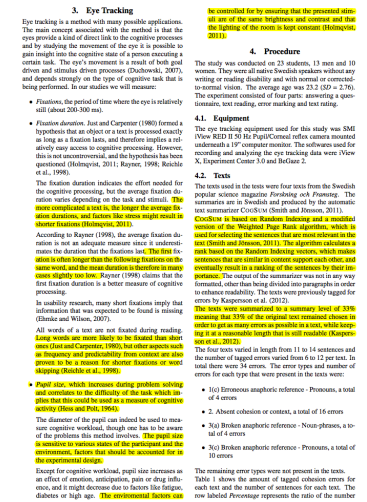
FODINA



Textanalys och textförenkling

- Vektorrums- och grafmodeller
 - Reducera den lingvistiska variationen och fånga semantiskt relaterade begrepp
- Bayesianska modeller
- Korpusinsamling
- Maskininlärning
- Utvärderingar
 - Guldstandarder
- Användarcentrerad utveckling

li.u LINKÖPINGS
UNIVERSITET



Textförenklingar

- Regelbaserat verktyg
 - Dependensgrammatik
- Modell för textförenkling
 - Korpus
 - Alla vanliga och lättlästa texter från offentliga förvaltningar i Sverige, myndigheter, kommuner, landsting m.fl
 - Parallellställning, alignment
 - Träna modell
 - LSTM, encoder-decoder

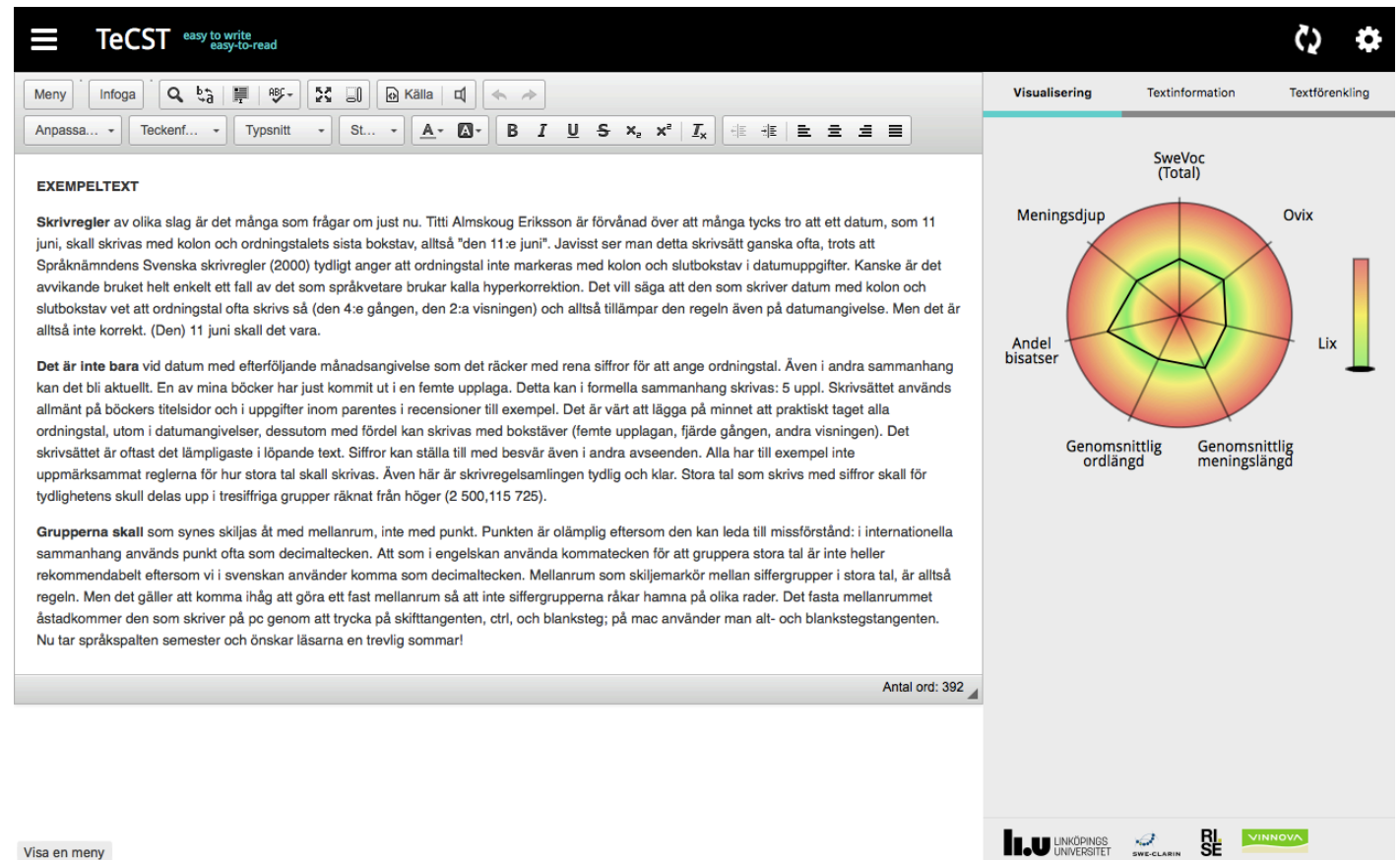
Textkomplexitetsmått

- Totalt 117 olika mått:
- Ytliga mått
 - Räknar ord och bokstäver
- Lexikala mått
 - Baserade på ordfrekvenser och grundläggande svensk vokablä, SweVoc
- Morfosyntaktiska mått
 - Bygger på en morfologisk analys av texten
- Syntaktiska mått
 - Egenskaper beräknade efter en syntaktisk analys av texten
- Textkvalitetsmått
 - Mått som vanligtvis används för att mäta läsbarhet

TeCST och SAPIS

<http://www.ida.liu.se/projects/scream/webapp/>

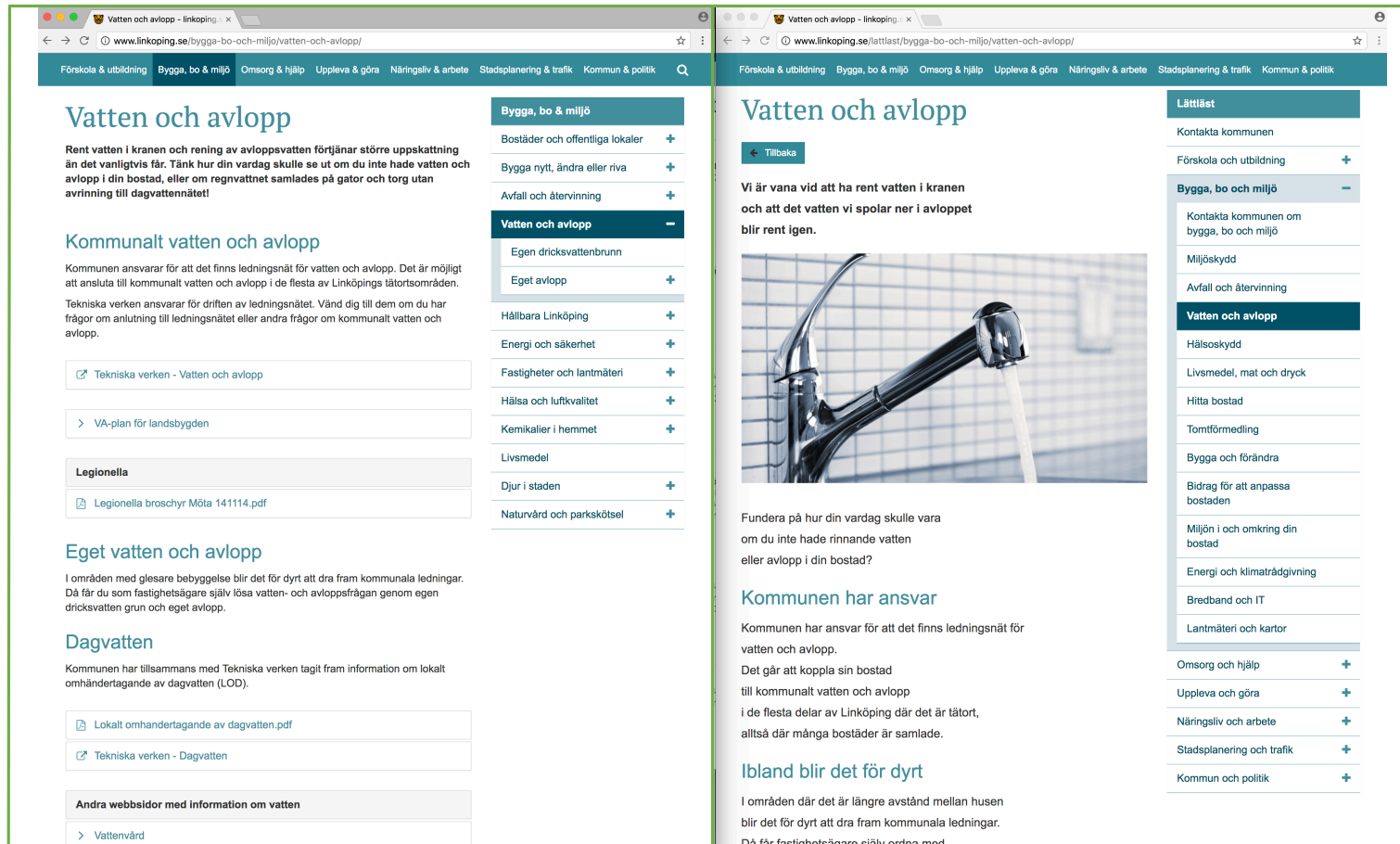
- Alla verktygen i en integrerad tjänst och ett REST API
- Visualisering av läsbarhetsmått



En korpus för att studera textförenkling

Evelina Rennes

Parallela texter



Monolingvistisk parallellställning

- Ordanvändning
- Grammatisk form
- Meningsordning

(---*---*---Lättläst---*---*---)
TITLE: Tekniska förvaltningen – Trelleborgs kommun

CONTENT: På tekniska förvaltningen arbetar man med att sköta gator och parker. Man planerar och gör kartor när det ska byggas. Förvaltningen har också hand om vatten och avlopp, sophämtning och energi.

(---*---*---Standard---*---*---)
TITLE: Tekniska förvaltningen – Trelleborgs kommun

CONTENT: Tekniska förvaltningens ansvar är att fullgöra tekniska nämndens uppgifter och ansvar för frågor i kommunen beträffande allmän platsmark, trafikens infrastruktur, våghållning, småbåtshamnarna, torghandel, avfallshantering, vattenförsörjning, avloppshantering, kommunens elnät samt kommunens markreserv. Tekniska förvaltningen finns i Rådhusets östra del, på Skyttsgatan 16 och Sjöviksvägen 6. Verksamheten bedrivs ute på anläggningar runt om i kommunen. Inom tekniska förvaltningen finns sju avdelningar: Administrativa avdelningen bistår samtliga avdelningar med stöd i bl.a. administration, personal, samordning, ekonomisk planering, budgetering, uppföljning och analys av förvaltningens olika verksamheter. Energiavdelningen har ansvar för elnät och distribution av el inom koncessionsområdet. Gatuvdelningen ansvarar för drift och underhåll av gator och torg, övriga allmänna platser, trafiksäkerhetsfrågor, salutorg, parkeringsövervakning, kommunens småbåtshamn, bidrag till enskilda vägar och kommunens gatubelysning. Parkavdelningen har hand om skötsel, förvaltning och nyanläggning av parkmark, lekplatser och utrustning i den gröna miljön. Projektavdelningen ansvarar för projektering, planering och genomförande av projekt inom förvaltningen. Avfalls- och återvinningsavdelningen hämtar hushållsavfall, vitvaror, trädgårds- och grovavfall samt slam från enskilda avlopp. Avdelningen ansvarar för driften av kommunens återvinningscentraler, återvinningsstationer och miljöstationer samt information till skolor och hushåll. VA-avdelningen ansvarar för kommunens vattenverk, avloppsreningsverk, pumpstationer, vattenledningar och avloppsledningar.

Korpus av svenska myndighetstexter

	Vanlig svenska	Lätt svenska
Antal dokument	115 027	2247
Antal meningar	1 333 474	26 461
Antal tecken	20 649 700	338 977
Meningslängd, medel	15	12
LIX	52	44
SweVoc	49%	50%

Parallellställning, alignment

- Meningslikhet baserad på likhet mellan word embeddings
 - Swectors (Fallgren et al., 2016)
- Tre metoder: Average alignment, Maximum alignment, Hungarian alignment (Song & Roth, 2015)
- Myndighetspar, inte dokumentpar

Song, Y. and Roth, D. (2015). Unsupervised sparse vector densification for short text similarity. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1275–1280.

Fallgren, P., Segeblad, J., and Kuhlmann, M. (2016). Towards a standard dataset of swedish word vectors. In *Proceedings of the Sixth Swedish Language Technology Conference (SLTC)*, Umeå, Sweden.

Alignmentmetoder

Method	Sentence pairs	Average sentence length*	
		Easy	Standard
AA	2543	7	6
MA	163,109	10	11
HA	1,907,130	10	15

* Indicated by the number of tokens.

Table 2: Description of the corpora generated by the different alignment methods.

Alignmentmetoder - Average

Method	Sentence pairs	Average sentence length*	
		Easy	Standard
AA	2543	7	6
MA	163,109	10	11
HA	1,907,130	10	15

* Indicated by the number of tokens.

Table 2: Description of the corpora generated by the different alignment methods.

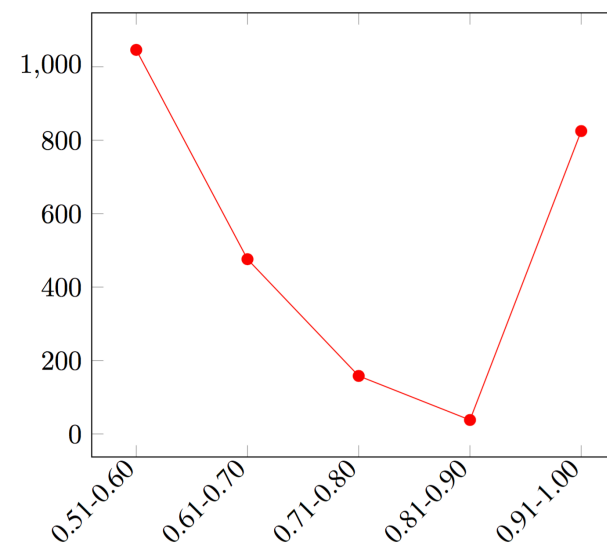


Figure 1: Number of aligned sentence pairs, for AA, within different similarity intervals.

Alignmentmetoder - Maximum

Method	Sentence pairs	Average sentence length*	
		Easy	Standard
AA	2543	7	6
MA	163,109	10	11
HA	1,907,130	10	15

* Indicated by the number of tokens.

Table 2: Description of the corpora generated by the different alignment methods.

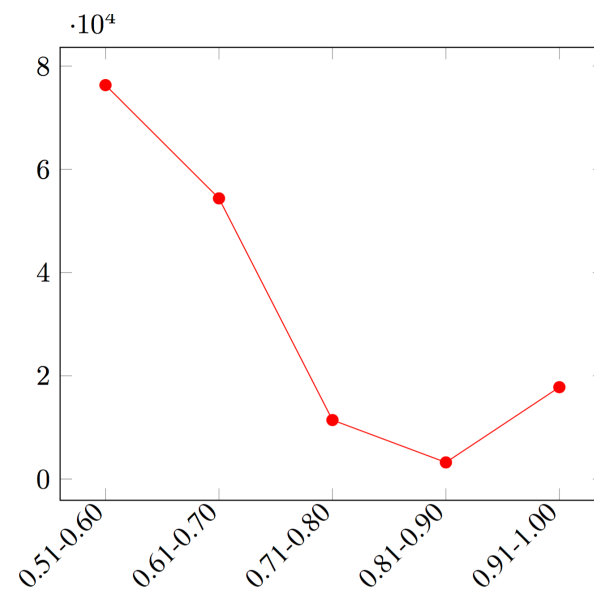


Figure 2: Number of aligned sentence pairs, for MA, within

Alignmentmetoder - Hungarian

Method	Sentence pairs	Average sentence length*	
		Easy	Standard
AA	2543	7	6
MA	163,109	10	11
HA	1,907,130	10	15

* Indicated by the number of tokens.

Table 2: Description of the corpora generated by the different alignment methods.

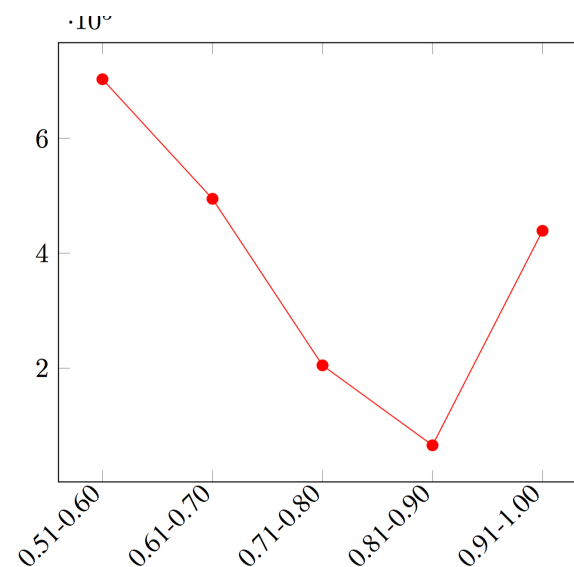


Figure 3: Number of aligned sentence pairs, for HA, within different similarity intervals.

Den är från 1700-talet.

Kycklingleden är en kvalitetsmärkt Värmlandsled

sim=0.9999999999999998

Arkenbiblioteket i Örnsköldsvik.

Örnsköldsvik från Lungviks-hållet.

sim=1.0

Fundera på hur din vardag skulle vara om du inte hade rinnande vatten eller avlopp i din bostad?

0.77

Tänk hur din vardag skulle se ut om du inte hade vatten och avlopp i din bostad, eller om regnvattnet samlades på gator och torg utan avrinning till dagvattennätet!

Kommunen har ansvar för att det finns ledningsnät för vatten och avlopp.

0.80

Kommunen ansvarar för att det finns ledningsnät för vatten och avlopp.

Det går att koppla sin bostad till kommunalt vatten och avlopp i de flesta delar av Linköping där det är tätort, alltså där många bostäder är samlade.

0.76

Det är möjligt att ansluta till kommunalt vatten och avlopp i de flesta av Linköpings tätortsområden.

I områden där det är längre avstånd mellan husen blir det för dyrt att dra fram kommunala ledningar.

0.79

I områden med glesare bebyggelse blir det för dyrt att dra fram kommunala ledningar.

Då får fastighetsägare själv ordna med rent vatten och avlopp.

0.53

Kommunen ansvarar för att det finns ledningsnät för vatten och avlopp.

0.72

Då får du som fastighetsägare själv lösa vatten- och avloppsfrågan genom egen dricksvatten grun och eget avlopp.

Inom den närmsta framtiden...

- Guldstandardutvärdering
- Småmodifikationer?
 - annat vektorset?
 - mer kontext/pos/ner?
- Utvärdering mha SMT

Corpus - IDA

www.ida.liu.se/~arnjo82/diginclude/corpus/

li.u

LINKÖPINGS
UNIVERSITET

IDA - Department of Computer and Information Science

Swedish web site

Search

Search IDA.LIU.se

Search

A — Z

LIU ▶ IDA ▶ arnjo82 ▶ Diginclude ▶ Corpus

Arne Jönsson

Research

Demo

Publications

Courses

Contact

Diginclude

Korpus

Vi har samlat in en korpus bestående av samtliga svenska myndigheters webbsidor såsom de såg ut i augusti 2017. Korpusen är uppdelad på de sidor som på något sätt kallas lättlästa respektive de andra sidorna. Totalt samlade vi in 2247 dokument (26461 meningar) på lättläst svenska och 115027 dokument (1333474 meningar) på vanlig svenska.

Dessa webbsidor har sedan parallellställts genom att för varje myndighet utgår från varje mening på lätt svenska och leta upp motsvarande meningar på standardsvenska. För denna parallellställning har vi använt tre olika algoritmer som tidigare använts för engelska (Song och Roth 2015), Average Alignment, Maximum Alignment samt Hungarian Alignment. Dessa tre parallellställda korpusar finns tillgängliga på denna sida.

2

Average alignment, totalt 2543 meningspar

2

Maximal alignment, totalt 163 109 meningspar

2

Hungarian alignment, totalt 1 907 130 meningspar

Song, Y. and Roth, D. (2015). Unsupervised sparse vector densification for short text similarity. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1275-1280

Officiell webbsida på RISE

INOMÅNS
TÄMPLAR

1377

Region
Östergötland

Linköping

Der ideer blir verklighet

FIDINA

CONVERTUS

SYMBOLBRUKET

briteback

The project is financed by Vinnova

Page responsible: Arne Jönsson

Last updated: 2017-10-09