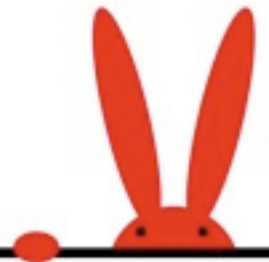


vad man som entreprenör i
språkteknologibranschen tycker
om språkresurser och språklig
infrastruktur

jussi karlgren

Stockholm, 2012

Gavagai



mål - varför språklig digital infrastruktur?

bättre utbildning
bättre forskning
mer livligt näringsliv
bättre tjänster för allmänheten

Copyright Gavagai

Gavagai



språkteknologi som industri

ingenjörsmässiga byggstenar - inte en tjänstesektor!

översättning

rättslingvistik

tal-till-text

text-till-tal

multimediasökning

digitala bibliotek och arkiv

omvärldsbevakning

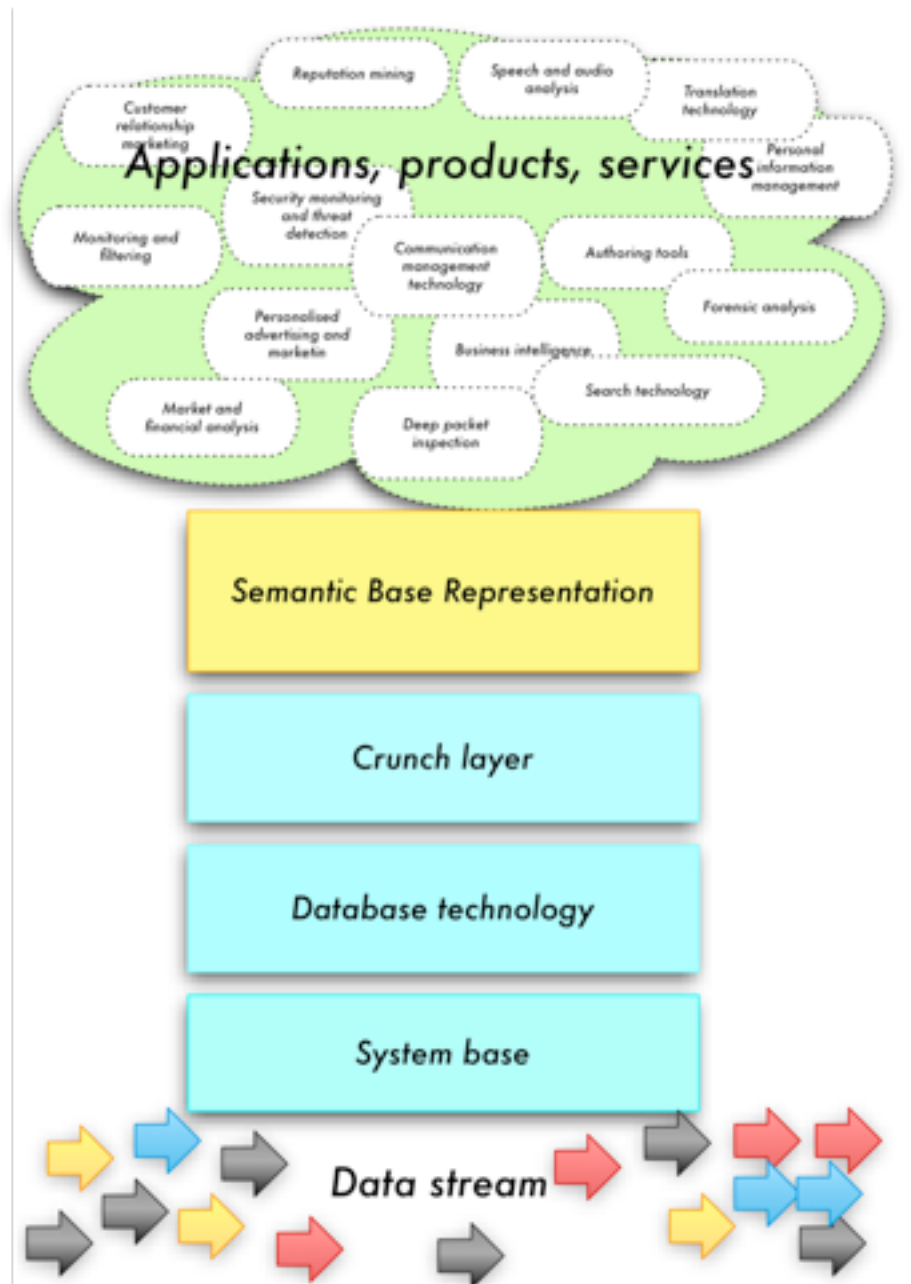
dialogsystem

är en motor även för innovation som görs av andra!

Gavagai



technology enabler: the big data stack



Gavagai



utmaningar

Copyright Gavagai

Gavagai



utmaningar

växande datamängder

Copyright Gavagai

Gavagai



utmaningar

växande datamängder

affärsmodeller och bristfällig fantasi

Copyright Gavagai

Gavagai



utmaningar

växande datamängder

affärsmodeller och bristfällig fantasi

beställarfeghet

Copyright Gavagai

Gavagai



utmaningar

växande datamängder

affärsmodeller och bristfällig fantasi

beställarfeghet

flerspråkighet och fragmentering

Copyright Gavagai

Gavagai



utmaningar

växande datamängder

affärsmodeller och bristfällig fantasi

beställarfeghet

flerspråkighet och fragmentering

låsta datasilon

Copyright Gavagai

Gavagai



resurser

Copyright 2009

Gavagai



resurser

forskningsfordon för informationsåtkomst sedan 50 år

Copyright Gavagai

Gavagai



resurser

forskningsfordon för informationsåtkomst sedan 50 år

fungerar fint om kopplat till uppgift

Copyright Gavagai

Gavagai



resurser

forskningsfordon för informationsåtkomst sedan 50 år

fungerar fint om kopplat till uppgift

blir hämsko om uppgiften fel eller obsolet

Copyright Gavagai

Gavagai



resurser

forskningsfordon för informationsåtkomst sedan 50 år

fungerar fint om kopplat till uppgift

blir hämsko om uppgiften fel eller obsolet

testuppgifter ett område statt i snabb omstöpning

Copyright Gavagai

Gavagai



resurser

forskningsfordon för informationsåtkomst sedan 50 år

fungerar fint om kopplat till uppgift

blir hämsko om uppgiften fel eller obsolet

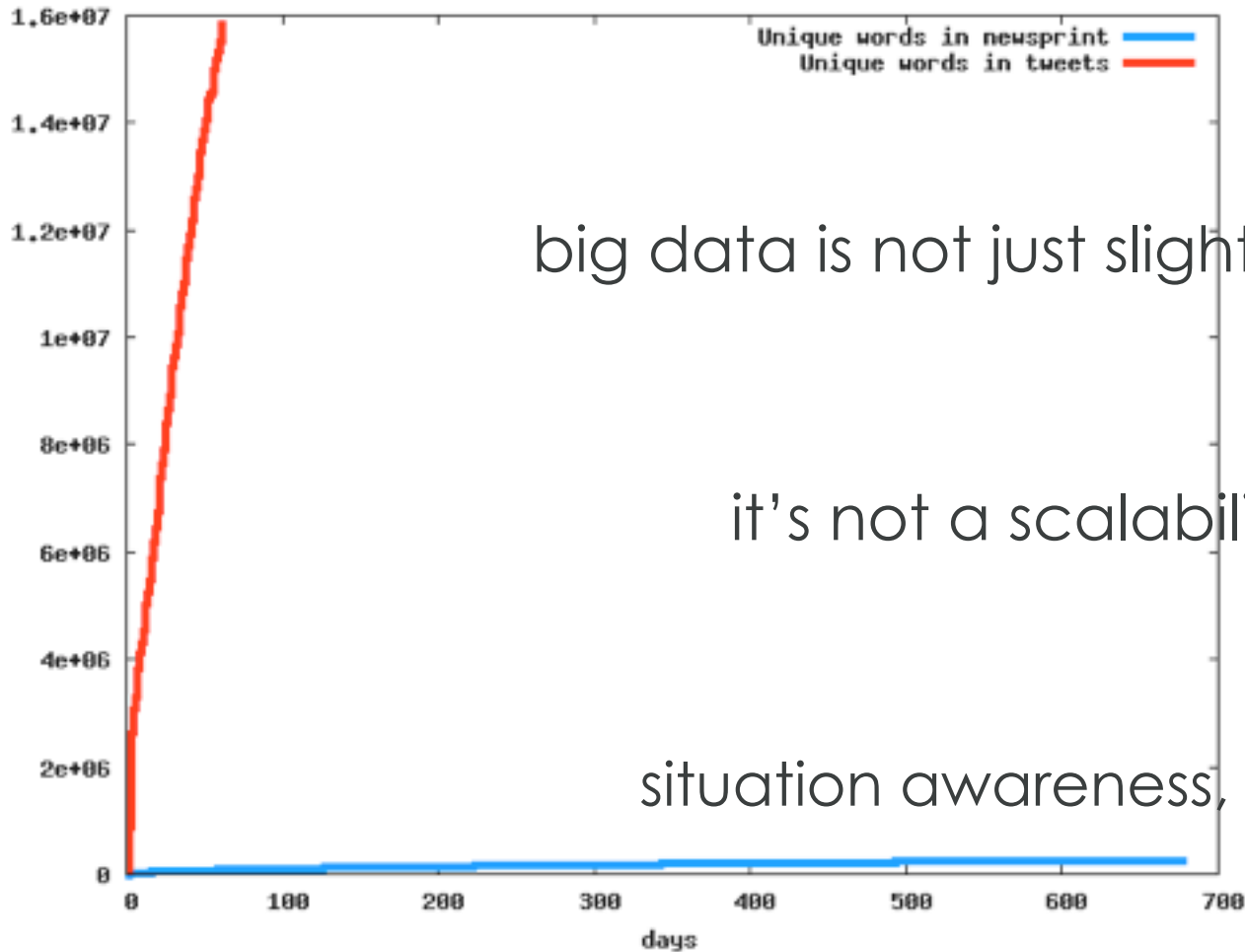
testuppgifter ett område statt i snabb omstöpning

jämför meteorologi!

Copyright Gavagai

Gavagai





big data is not just slightly larger data

it's not a scalability issue

situation awareness, not search



prisnivå för data

Trendiction - Social media and news in many languages:

News 5000EUR/month; Blogs 4500EUR/month; Fora 8000EUR/month

Moreover - Social media and news in many languages:

News: \$3200/month; Blogs and fora: \$3200/month

Spinn3r - The largest aggregator of English social media:

Blogs: \$5500/month; News: \$2500/month

Gnip - Twitter, Google+, Facebook, Tumblr, Wordpress, Disqus:

\$3000/month for six sources

Omgili (now part of Buzzilla) - A forum search engine with firehose access:

\$3500/month

Bloomberg - Financial data:

\$3000/month.

Copyright Gavagai

Gavagai



SÅ vad behöver vi?

Copyright Gavagai

Gavagai



SÅ vad behöver vi?

rådata med full täckning

Copyright Gavagai

Gavagai



SÅ vad behöver vi?

rådata med full täckning
på de språk som används i Sverige

Copyright Gavagai

Gavagai



SÅ vad behöver vi?

rådata med full täckning
på de språk som används i Sverige
oförfinat, oanalyserat

Gavagai Sverige

Gavagai



SÅ vad behöver vi?

rådata med full täckning
på de språk som används i Sverige
oförfinat, oanalyserat

kompetent personal

Gavagai Sverige

Gavagai



SÅ vad behöver vi?

rådata med full täckning
på de språk som används i Sverige
oförfinat, oanalyserat

kompetent personal

öppna standarder (för tillträde till marknader)

Gavagai Sverige

Gavagai



SÅ vad behöver vi?

rådata med full täckning
på de språk som används i Sverige
oförfinat, oanalyserat

kompetent personal

öppna standarder (för tillträde till marknader)

beställarkompetens och beställarsjälvkänsla

Gavagai



vad vi inte vill ha

små samlingar av teori-specifika experimentresultat
(trädbanker ger trädbanksforskning)

vinstdrivande subventionerade mellanhänder

programkod ("vad är en tagg"?)

Copyright Gavagai

Gavagai



varför inte programkod?

Copyright 2006

Gavagai



varför inte programkod?

för tillämpning måste programkod vara tiptop

Copyright Gavagai

Gavagai



varför inte programkod?

för tillämpning måste programkod vara tiptop

bäst programkod kommer ur
gemensamhetssatsningar

Copyright Gavagai

Gavagai



varför inte programkod?

för tillämpning måste programkod vara tiptop

bäst programkod kommer ur
gemensamhetssatsningar

en plattform för att understödja sådant är viktigare
än själva koden

Copyright Gavagai

Gavagai



varför inte programkod?

för tillämpning måste programkod vara tiptop

bäst programkod kommer ur
gemensamhetssatsningar

en plattform för att understödja sådant är viktigare
än själva koden

(innovationsaktiviteter innovationsgemenskaper!)

Copyright Gavagai

Gavagai



frågor till den som planerar infrastruktur

vilket eller vilka är användningsfallen?

vilka är idag flaskhalsarna?

hur ser ni på de organisatoriska ramarna?

Copyright Gavagai

Gavagai

