NewsReader

recording history
by processing massive streams of daily news

VU UNIVERSITY AMSTERDAM

ScraperWiki

Universidad del País Vasco
Euskal Herriko Unibertsitatea

LexisNexis

FONDAZIONE BRUNO KESSLER

synerscope
connecting the dots

NewsReader
POST HOC ERGO PROPTER HOC

COOPERATION

HOW DID THE WORLD CHANGE YESTERDAY?

Can we handle the news?

- Information broker LexisNexis archives:
  - 1.5 millions news articles on a single working day
  - 30,000 different sources

How did the Car industry change during the financial crisis?

- 6 million English articles on the car industry in the LexisNexis archive for the last 10 years

- 2 million Google hits for “Volkswagen takeover” not sorted by publication date

Google

piek vossen 0 + Share

Trends

Web Search Interest: **volkswagen**. Worldwide, 2004 - present.

Explore trends

Hot searches

Interest over time

The number 100 represents the peak search interest

News headlines Forecast

Search terms

volkswagen

+ Add term

Other comparisons

Limit to

THE PROBLEM

10k entities per year

200k mentions per year

HOW MANY CHANGES?

6 MILLION ARTICLES

1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11 12 13 14 2015

Past *New* *Past* *New* *Speculation*

On 16 September 2008, Porsche *increased its shares* by another 4.89%, in effect *taking control of* the company, with more than 35% of the voting rights.

6 Jan 2009 – Porsche has been on *a quest to takeover* VW for more than two years.

DAILY NEWS TSUNAMI

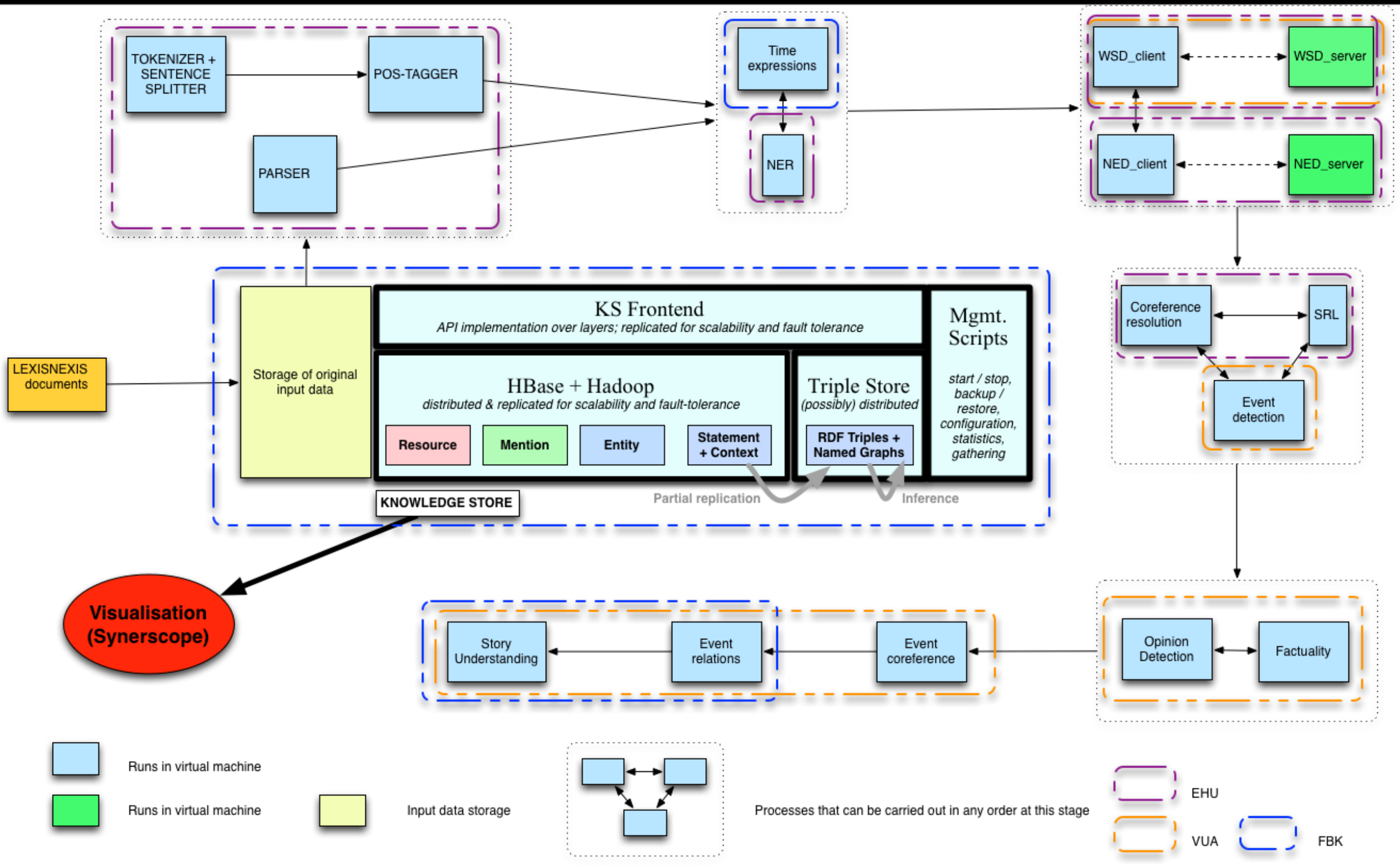
- Volume is too big: 1,5 million items each working day
- Repeated and duplicated: we cannot distinguish new from old
- Incomplete and piecemeal: we need to read all to get a complete picture
- Actual and speculated events: we cannot distinguish the realis from irrealis (speculations, fears and hopes)
- Inconsistent and contradictory: we cannot tell true from false (who to believe)
- Opinionated and selective: we do not realize the bias of our sources

What if computers could read the news?

NewsReader (ict316404)

- ***Reading Technology*** to process massive streams of news from many different sources in 4 languages (English, Dutch, Spanish and Italian):
  - Recording the *changes* in the world as they are told in the media over long periods of time → ***history-recorder***.
  - ***What*** happened, ***where*** and ***when***, ***who*** was involved.
  - What ***temporal*** and ***causal*** relations hold.
  - Who made what statement, where do sources agree and disagree: ***provenance***!
  - KnowledgeStore that handles ***dynamic growth*** of information, reflecting long-term developments.
  - ***Visualize*** massive amount of changes as *stories*, *scripts*, *plots* to provide efficient access

System Architecture



General approach

- Representation centered architecture
  - Predefined NLP Annotation Format (NAF)
  - Layered annotation format
  - All modules inputstream = NAF, outputstream = NAF with new layer
  - Easy to add new layers on top of given layers
  - Design alternative pipelines by combining modules

Grounded annotation framework

- **GAF**: groundedannotationformat.org.
- Distinguishes between mentions of entities and events in sources (text, images, movies, databases, sensors) and the representation of instances in the assumed world.
- Mentions are semantically represented in the **NAF** (NLP Annotation Format) representation of the text (same instances mentioned in at different places in the text and in different texts).
- Instances are semantically represented in **SEM** (RDF-based Simple Event Model, Van Hage et al 2011) using URIs.
- gaf:denotes and gaf:denotedBy links to connect the two (**PROV-O**)

NLP Annotation Format (NAF)

- Represents linguistic annotations.
  - Stand-o, multi-layered annotation format.
  - Based on XML.
- Compatible with main standards
  - LAF, Ide et al., 2003
  - GATE (Cunningham et al., 1996)
  - UIMA, (Ferrucci and Lally, 2004)
  - : : :
- Allows parallel processing.
- Can be exported to RDF triplets.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps> ... </deps>
  <entities> ... </entities>
  <coreferences> ... </coreferences>
  <srl> ... </srl>
  <factualitylayer> ... </factualitylayer>
</NAF>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader>
    <fileDesc creationtime="2013-01-01"/>
    <public publicId="57D5-K8H1-JCBN-04H0" uri="2013/1/1/57D5-K8H1-JCBN-04H0.xml"/>
    <linguisticProcessors layer="text">
      <lp name="ixa-pipe-tok-en" timestamp="2013-11-26 14:06:46" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="srl">
      <lp name="ixa-pipe-srl-en" timestamp="2013-11-26 14:07:39" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="factuality">
      <lp name="vua-factuality" timestamp="2013-11-26T14:07:44Z" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="terms">
      <lp name="ixa-pipe-pos-en" timestamp="2013-11-26 14:06:46" version="1.0"/>
      <lp name="vua-multiword-tagger" timestamp="2013-11-26 14:06:51" version="1.0"/>
      <lp name="VUA-DSC-WSD" timestamp="2013-11-26T14:06:55CET" version="8nov2013_v1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="coreferences">
      <lp name="vua-event-coref-intradoc-lemma-baseline" timestamp="2013-11-26 14:07:43" version="1.0"/>
      <lp name="vua-entity-coref-intradoc-reference-baseline" timestamp="2013-12-13 17:23:08" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="deps">
      <lp name="ixa-pipe-srl-en" timestamp="2013-11-26 14:07:39" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="opinions">
      <lp name="VUA opinion miner. CRF deluxe" timestamp="2013-11-26T14:06:54CET" version="8nov2013_1.1"/>
    </linguisticProcessors>
    <linguisticProcessors layer="timex3">
      <lp name="TimePro" timestamp="2013-11-26 14:07:42.631" version="2.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="entities">
      <lp name="ixa-pipe-nerc-en" timestamp="2013-11-26 14:06:53" version="1.0"/>
    </linguisticProcessors>
    <linguisticProcessors layer="ned">
      <lp name="ixa-pipe-spotlight" timestamp="2013-11-26 14:06:56" version="1.0"/>
    </linguisticProcessors>
  </nafHeader>
  <raw>Toyota starts remodelled Crown sales


Toyota Motor has begun selling a redesigned Crown, its oldest sedan still in production, today in Japan in a bid to boost sales that have

The 14th-generation Crown starts at à¸…3.53 million (Bt1.2 million), the company, Asia's biggest carmaker, said in a statement. The sedan
Toyota targets deliveries of 4,000 of the sedans a month in Japan, according to the statement, compared with sales averaging 17,000 a mont
_"To attract people to the car today, we had to redesign it."_president Akio Toyoda told reporters today in Tokyo.
The Crown remained Toyota_s flagship luxury car in Japan after it introduced the Lexus in the US in 1989 to compete with Daimler's Mercede
Among Japanese seniors, the Crown still symbolises luxury, and still carries the image as the CEO's car," said Toshihiro Nagahama, chief e
Bloomberg</raw>
  <text>
    <wf id="w1" length="6" offset="0" sent="1">Toyota</wf>
    <wf id="w2" length="6" offset="7" sent="1">starts</wf>
    <wf id="w3" length="10" offset="14" sent="1">remodelled</wf>
    <wf id="w4" length="5" offset="25" sent="1">Crown</wf>
    <wf id="w5" length="5" offset="31" sent="1">sales</wf>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms>
    <term id="t1" lemma="Toyota" morphofeat="NNP" pos="R" type="close">
      <span>
        <!--Toyota-->
        <target id="w1"/>
      </span>
    </term>
    <term id="t2" lemma="start" morphofeat="VBZ" pos="V" type="open">
      <span>
        <!--starts-->
        <target id="w2"/>
      </span>
    </term>
    <term id="t3" lemma="remodel" morphofeat="VBN" pos="V" type="open">
      <span>
        <!--remodelled-->
        <target id="w3"/>
      </span>
    </term>
    <term id="t4" lemma="Crown" morphofeat="NNP" pos="R" type="close">
      <span>
        <!--Crown-->
        <target id="w4"/>
      </span>
    </term>
    <term id="t5" lemma="sale" morphofeat="NNS" pos="N" type="open">
      <span>
        <!--sales-->
        <target id="w5"/>
      </span>
    </term>
    <term id="t6" lemma="Toyota" morphofeat="NNP" pos="R" type="close">
      <span>
        <!--Toyota-->
        <target id="w6"/>
      </span>
    </term>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps>
    <!--SBJ(starts,Toyota)-->
    <dep from="t2" rfunc="SBJ" to="t1"/>
    <!--NMOD(sales,remodelled)-->
    <dep from="t5" rfunc="NMOD" to="t3"/>
    <!--NMOD(sales,Crown)-->
    <dep from="t5" rfunc="NMOD" to="t4"/>
    <!--NMOD(Motor,sales)-->
    <dep from="t7" rfunc="NMOD" to="t5"/>
    <!--NMOD(Motor,Toyota)-->
    <dep from="t7" rfunc="NMOD" to="t6"/>
    <!--SBJ(has,Motor)-->
    <dep from="t8" rfunc="SBJ" to="t7"/>
    <!--OBJ(starts,has)-->
    <dep from="t2" rfunc="OBJ" to="t8"/>
    <!--VC(has,begun)-->
    <dep from="t8" rfunc="VC" to="t9"/>
    <!--OPRD(begun,selling)-->
    <dep from="t9" rfunc="OPRD" to="t10"/>
    <!--NMOD(sedan,a)-->
    <dep from="t17" rfunc="NMOD" to="t11"/>
    <!--NMOD(sedan,redesigned)-->
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps> ... </deps>
  <entities>
    <entity id="e1" type="person">
      <references>
        <span>
          <!--Toyota Motor-->
          <target id="t6"/>
          <target id="t7"/>
        </span>
      </references>
      <externalReferences>
        <externalRef reference="http://dbpedia.org/resource/Toyota" resource="spotlight_v1"/>
      </externalReferences>
    </entity>
    <entity id="e2" type="location">
      <references>
        <span>
          <!--Crown-->
          <target id="t13"/>
        </span>
      </references>
      <externalReferences>
        <externalRef reference="http://dbpedia.org/resource/The_Crown" resource="spotlight_v1"/>
      </externalReferences>
    </entity>
    <entity id="e3" type="location">
      <references>
        <span>
          <!--Japan-->
          <target id="t24"/>
        </span>
      </references>
      <externalReferences>
        <externalRef reference="http://dbpedia.org/resource/Japan" resource="spotlight_v1"/>
      </externalReferences>
    </entity>
    <entity id="e4" type="location">
      <references>
        <span>
          <!--Asia-->
          <target id="t63"/>
        </span>
      </references>
      <externalReferences>
        <externalRef reference="http://dbpedia.org/resource/Asia" resource="spotlight_v1"/>
      </externalReferences>
    </entity>
```



```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps> ... </deps>
  <entities> ... </entities>
  <coreferences>
    <coref id="coe26" type="event">
      <span>
        <target id="t159"/>
      </span>
    </coref>
    <coref id="coe17" type="event">
      <span>
        <target id="t90"/>
      </span>
    </coref>
    <coref id="coe14" type="event">
      <span>
        <target id="t75"/>
      </span>
    </coref>
    <coref id="coe29" type="event">
      <span>
        <target id="t181"/>
      </span>
    </coref>
    <coref id="coe22" type="event">
      <span>
        <target id="t129"/>
      </span>
    </coref>
    <coref id="coe31" type="event">
      <span>
        <target id="t199"/>
      </span>
    </coref>
    <coref id="coe30" type="event">
      <span>
        <target id="t190"/>
      </span>
    </coref>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps> ... </deps>
  <entities> ... </entities>
  <coreferences> ... </coreferences>
  <srl>
    <predicate id="pr1">
      <!--starts-->
      <externalReferences>
        <externalRef reference="start.01" resource="PropBank"/>
        <externalRef reference="begin-55.1" resource="VerbNet"/>
        <externalRef reference="begin-55.1-1" resource="VerbNet"/>
        <externalRef reference="Activity_start" resource="FrameNet"/>
        <externalRef reference="Process_start" resource="FrameNet"/>
        <externalRef reference="Setting_fire" resource="FrameNet"/>
        <externalRef reference="grammatical" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t2"/>
      </span>
      <role id="rl1" semRole="A0">
        <!--Toyota-->
        <externalReferences>
          <externalRef reference="begin-55.1#Agent" resource="VerbNet
          <externalRef reference="Activity_start#Agent" resource="Fra
        </externalReferences>
        <span>
          <target head="yes" id="t1"/>
        </span>
      </role>
      <role id="rl2" semRole="A1">
        <!--remodelled Crown sales Toyota Motor has begun selling a r
        <externalReferences>
          <externalRef reference="begin-55.1#Theme" resource="VerbNet
          <externalRef reference="Activity_start#Activity" resource="
          <externalRef reference="Process_start#Event" resource="Fram
        </externalReferences>
        <span>
          <target id="t3"/>
          <target id="t4"/>
          <target id="t5"/>
          <target id="t6"/>
          <target id="t7"/>
          <target head="yes" id="t8"/>
          <target id="t9"/>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<NAF version="v3" xml:lang="en">
  <nafHeader> ... </nafHeader>
  <raw> ... </raw>
  <text> ... </text>
  <terms> ... </terms>
  <deps> ... </deps>
  <entities> ... </entities>
  <coreferences> ... </coreferences>
  <srl> ... </srl>
  <factualitylayer>
    <factvalue confidence="0.7668557322515226" id="w174" prediction="CT+"/>
    <factvalue confidence="0.6780538614109628" id="w199" prediction="CT+"/>
    <factvalue confidence="0.5918886049283905" id="w190" prediction="Uu"/>
    <factvalue confidence="0.7447324223480208" id="w122" prediction="CT+"/>
    <factvalue confidence="0.9419688057124761" id="w105" prediction="CT+"/>
    <factvalue confidence="0.8623865274452669" id="w152" prediction="Uu"/>
    <factvalue confidence="0.9454953981216198" id="w243" prediction="CT+"/>
    <factvalue confidence="0.8988616864127015" id="w113" prediction="CT+"/>
    <factvalue confidence="0.8110307066725162" id="w3" prediction="CT+"/>
    <factvalue confidence="0.9012014206275009" id="w29" prediction="Uu"/>
    <factvalue confidence="0.8511802384853364" id="w253" prediction="CT+"/>
    <factvalue confidence="0.5980360928428445" id="w33" prediction="Uu"/>
    <factvalue confidence="0.694483814663816" id="w32" prediction="Uu"/>
    <factvalue confidence="0.9597434500690043" id="w162" prediction="CT+"/>
    <factvalue confidence="0.7941969984929049" id="w212" prediction="CT+"/>
    <factvalue confidence="0.7194074186850347" id="w144" prediction="Uu"/>
    <factvalue confidence="0.8222445666730973" id="w2" prediction="CT+"/>
    <factvalue confidence="0.49102167172847705" id="w8" prediction="CT+"/>
    <factvalue confidence="0.5243743897158392" id="w10" prediction="CT+"/>
    <factvalue confidence="0.845733066248451" id="w93" prediction="CT+"/>
    <factvalue confidence="0.7396445230013207" id="w123" prediction="CT+"/>
    <factvalue confidence="0.664534108915098" id="w68" prediction="CT+"/>
    <factvalue confidence="0.5037103625039135" id="w9" prediction="CT+"/>
    <factvalue confidence="0.7560929572382062" id="w208" prediction="CT+"/>
    <factvalue confidence="0.5434615557832647" id="w12" prediction="CT+"/>
    <factvalue confidence="0.8708945295334155" id="w181" prediction="CT+"/>
    <factvalue confidence="0.895767186793211" id="w81" prediction="CT+"/>
    <factvalue confidence="0.8888553061579482" id="w154" prediction="Uu"/>
    <factvalue confidence="0.6304588396907782" id="w170" prediction="CT+"/>
    <factvalue confidence="0.8409174047710798" id="w238" prediction="CT+"/>
    <factvalue confidence="0.9032806950281298" id="w110" prediction="CT+"/>
    <factvalue confidence="0.7378353260791006" id="w224" prediction="Uu"/>
    <factvalue confidence="0.8373826853321445" id="w75" prediction="CT+"/>
  </factualitylayer>
</NAF>
```

Simple Event Model (SEM)

- Models events and participants
- who did what, when and where.
- Derived from various sources
- multiple docs, images, sensory data, ...
- Represents partial and contradictory information.
- Includes basic relations between events: subEventOf, causes.
- Final output of pipeline is representing following SEM+
- aggregated representation of events.

MENTIONS → INSTANCES → MENTIONS

Forbes
4/23/2004 @ 5:01PM
http://www.forbes.com/2004/04/23/cz_jf_0423flint.html

DaimlerChrysler just refused to make a $7 billion to $8 billion cash infusion to the floundering company Mitsubishi . ... His tactics led to massive investments in American Chrysler (a takeover), in Mitsubishi 37% ownership and control) and Korean Hyundai (10% and no control

New Zealand Herald,
Monday Apr 26, 2004
http://www.nzherald.co.nz

Schrempp may have suffered his own personal Waterloo on Friday when Daimler's board voted to pull the plug on troubled Japanese carmaker Mitsubishi Motors rather than pump in billions of euros to keep the company on financial life support.

WHAT: decision
WHO: DaimlerChrysler
WHEN: Friday, April, 23, 2004

NOT

WHAT: invest
WHO: DaimlerChrysler
WHO: Mitsubishi
WHO: $7-8 billion euros

New York Times, By MARK LANDLER
Published: April 24, 2004
http://www.nytimes.com/2004/04/24/

Even Mr. Schrempp's hold on the chief executive's job at DaimlerChrysler seems shaky in the wake of his company's unexpected refusal to aid a bailout of the financially troubled Mitsubishi

Automotive News, Monday
Apr 26, 2004:3
http://www.autonews.com

The decision not to bail out Mitsubishi Motors Corp raises fresh doubts about the future of DaimlerChrysler CEO Juergen Schrempp

NAF example

Toyota brought Lexus to Japan in 2005.

```
<predicate id="pr36">
  <!--brought-->
  <externalReferences>
    <externalRef reference="bring.01" resource="PropBank"/>
    <externalRef reference="bring-11.3-1" resource="VerbNet"/>
    <externalRef reference="Bringing" resource="FrameNet"/>
  </externalReferences>
  <span><target id="t199"/></span>
  <role id="rl84" semRole="A0">
    <!--Toyota-->
    <externalReferences>
      <externalRef reference="bring-11.3#Agent" resource="VerbNet"/>
    </externalReferences>
    <span><target head="yes" id="t198"/></span>
  </role>
  <role id="rl85" semRole="A1">
    <!--Lexus-->
    <externalReferences>
      <externalRef reference="bring-11.3#Theme" resource="VerbNet"/>
    </externalReferences>
    <span><target head="yes" id="t200"/></span>
  </role>
  <role id="rl86" semRole="A3">
    <!--to Japan-->
    <span><target head="yes" id="t201"/><target id="t202"/>
    </span>
  </role>
  <role id="rl87" semRole="AM-TMP">
    <!--in 2005-->
    <span><target head="yes" id="t203"/><target id="t204"/>
    </span>
  </role>
</predicate>
```

```
<entities>
  <entity id="e1" type="person">
    <references>
      <span>
        <!--Toyota Motor-->
        <target id="t6"/>
        <target id="t7"/>
      </span>
    </references>
    <externalReferences>
      <externalRef reference="http://dbpedia.org
      /resource/Toyota" resource="spotlight_v1"/>
    </externalReferences>
  </entity>
```

```
<coref id="coentity1" type="person">
  <span>
    <!--Toyota motor-->
    <target id="t6"/>
    <target id="t7"/>
  </span>
  <span>
    <!--Toyota-->
    <target id="t198"/>
  </span>
</coref>
```

SEM in TriG format

EVENT INSTANCE

<nwr:data/cars/2013/1/1/5758-BPN1-F0J6-D2T2.xml#sellEvent>

a sem:Event , nwr:contextual , fn:Commerce_sell ;

rdfs:label "sell" ;

gaf:denotedBy

<nwr:data/cars/2013/1/1/5758-BPN1-F0J6-D2T2.xml#char=1352,1356&word=w251&term=t251> ,

<nwr:data/cars/2013/1/1/5760-PM51-JD34-P4H7.xml#char=1536,1540&word=w275&term=t275>.

SEM in TriG format

ENTITY INSTANCE

<http://dbpedia.org/resource/Toyota>

a sem:Actor , nwr:person , nwr:organization ,
nwr:framenet/Commerce_sell#Seller> ;

rdfs:label "Toyota" , "Toyota motor" ;

gaf:denotedBy

<nwr:data/cars/2013/1/1/5760-PM51-JD34-P4RM.xml#char=98,104&word=w18&term=t18> ,

<nwr:data/cars/2013/1/1/57K5-FKK1-DYBW-2534.xml#char=44934,44940&word=w8114&term=t8114> .

Semantic relations as named graphs

<nwr:/data/cars/2013/1/1/5758-BPN1-F0J6-D2T2.xml#pr25,rl55> {
  <nwr:data/cars/2013/1/1/5722-S821-F0J6-D48N.xml#sellEvent>
    sem:hasActor <http://dbpedia.org/resource/Magyar_Suzuki> .
}
<nwr:data/cars/2013/1/1/5760-PM51-JD34-P4H7.xml#pr46,rl114> {
  <nwr:data/cars/2013/1/1/5758-BPN1-F0J6-D2T2.xml#sellEvent>
    sem:hasPlace <http://dbpedia.org/resource/South_Africa> .
}
<nwr:data/cars/2013/1/1/5760-PM51-JD34-P4H7.xml#docTime_26> {
  <nwr:data/cars/2013/1/1/5760-PM51-JD34-P4H7.xml#sellEvent>
    sem:hasTime <nwr:time/2013-01-01> .
}

Properties of relations

PROVENANCE

<nwr:data/cars/2013/1/1/57R8-5451-F0J6-D2GH.xml#pr25,rl55>

gaf:denotedBy

<nwr:data/cars/2013/1/1/57R8-5451-F0J6-D2GH.xml#rl55> ;

prov-o:wasAttributedTo

<nwr:sourceowner/Peru_Autos_Report> .

FACTUALITY

<nwr:data/cars/2013/1/1/57K5-FKK1-DYBW-2534.xml#facValue_1125>

{

<nwr:data/cars/2013/1/1/57K5-FKK1-DYBW-2534.xml#sellEvent>

nwr:hasFactBankValue

"CT+" .}

Cross-document event coreference

- Instance based event-coreference:
  - All event mentions with same lemma and same time anchor
    - Share at least one actor (possibly DBPedia URI)
    - Share at least one place (possibly DBPedia URI)
- Aggregation of SEM instances from NAF mentions and the extraction of provenance layers through named graphs
- http://ic.vupr.nl/~ruben/vua-eventcoreference.ttl/

ON A WORKING DAY

Friday
10:15
February 6

Tuesday
09:34
February 10

1 MILLION ARCTICLES
FROM 30,000 SOURCES

1 ARTICLE

text

text

text

100 EVENT MENTIONS PER ARTICLE

naf

naf

naf

5 EVENT INSTANCE PER ARTICLE

sem

sem

Knowledge Store

DBpedia

PLACES

EVENTS

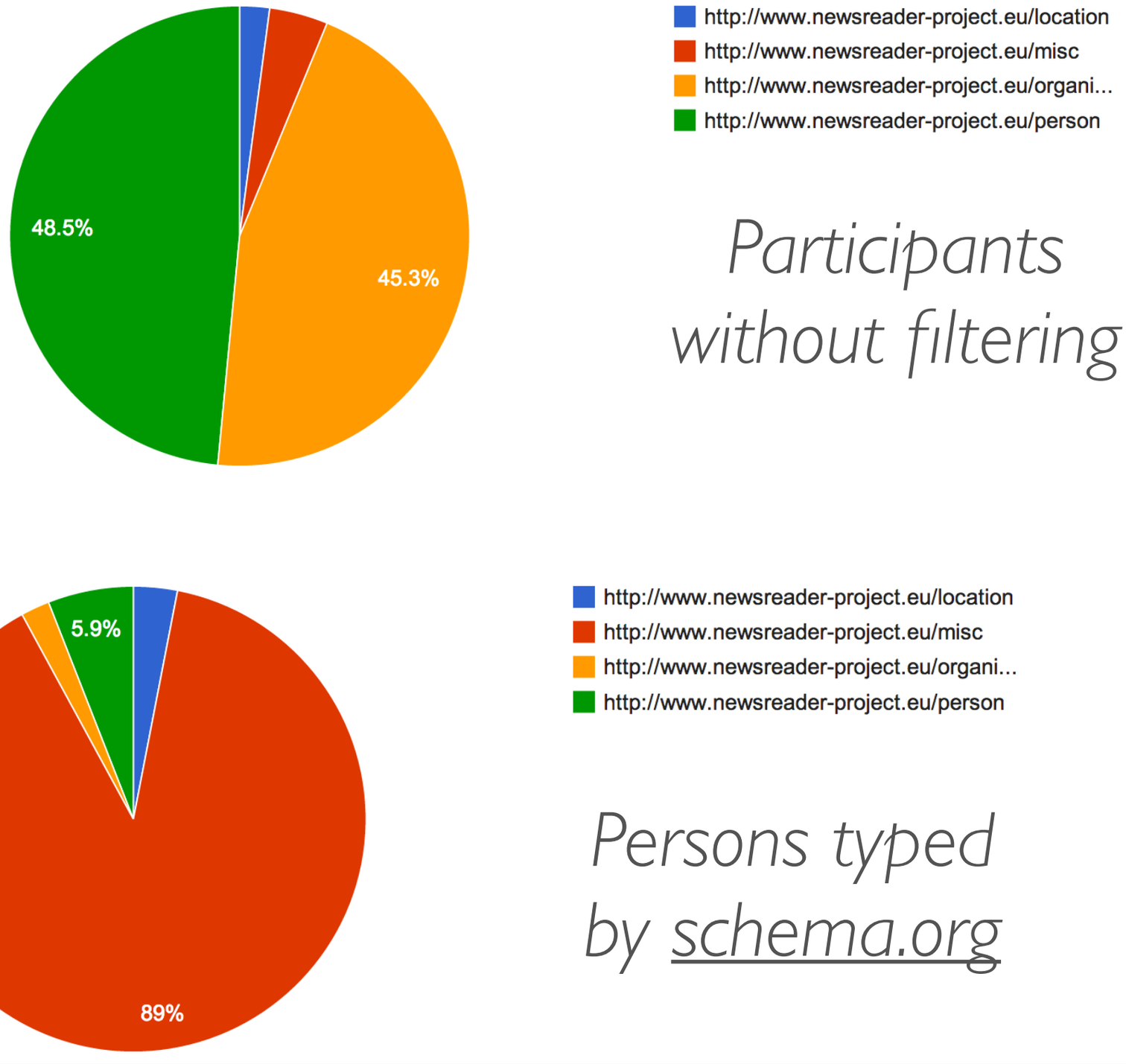
ENTITIES

1995 96 97 98 99 2000 01 02 03 04 2005 06 07 08 09 2010 11 12 13 14 2015

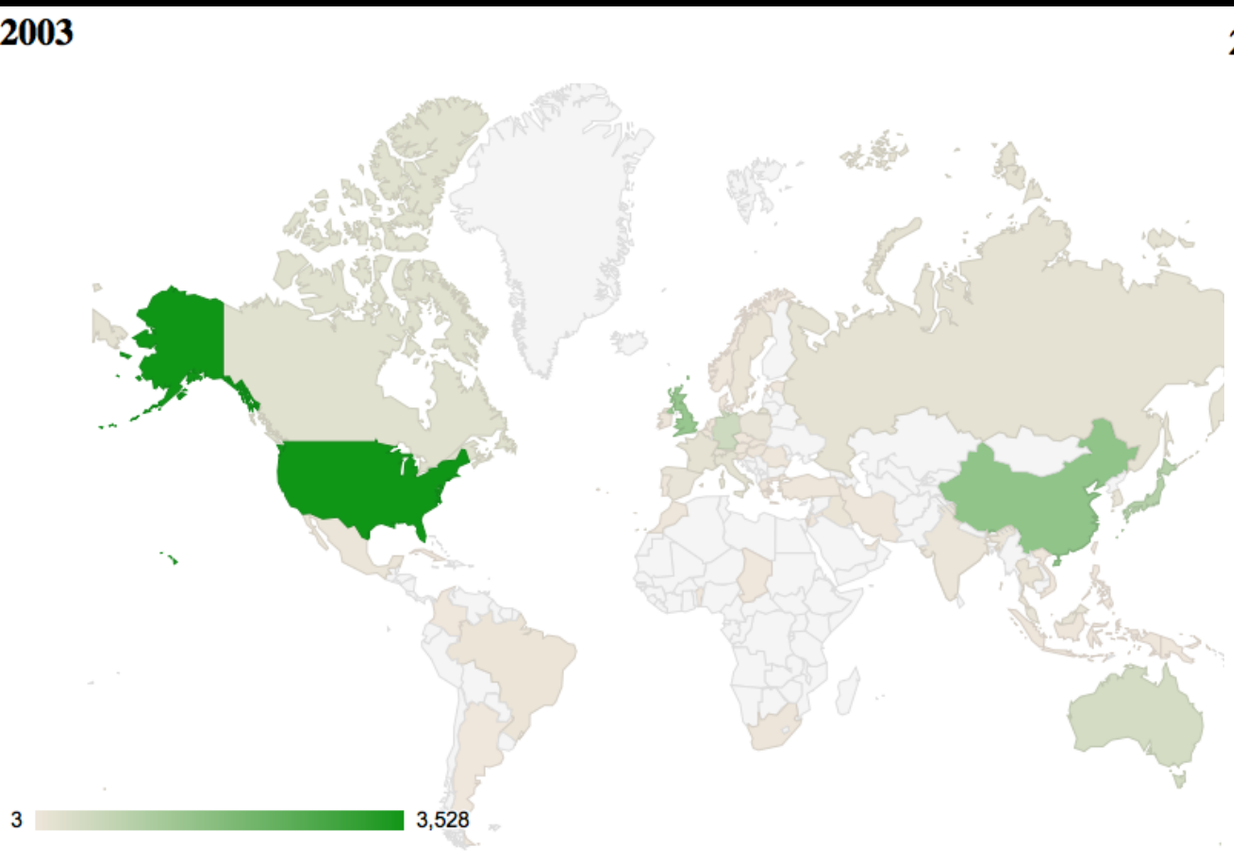
Available libraries

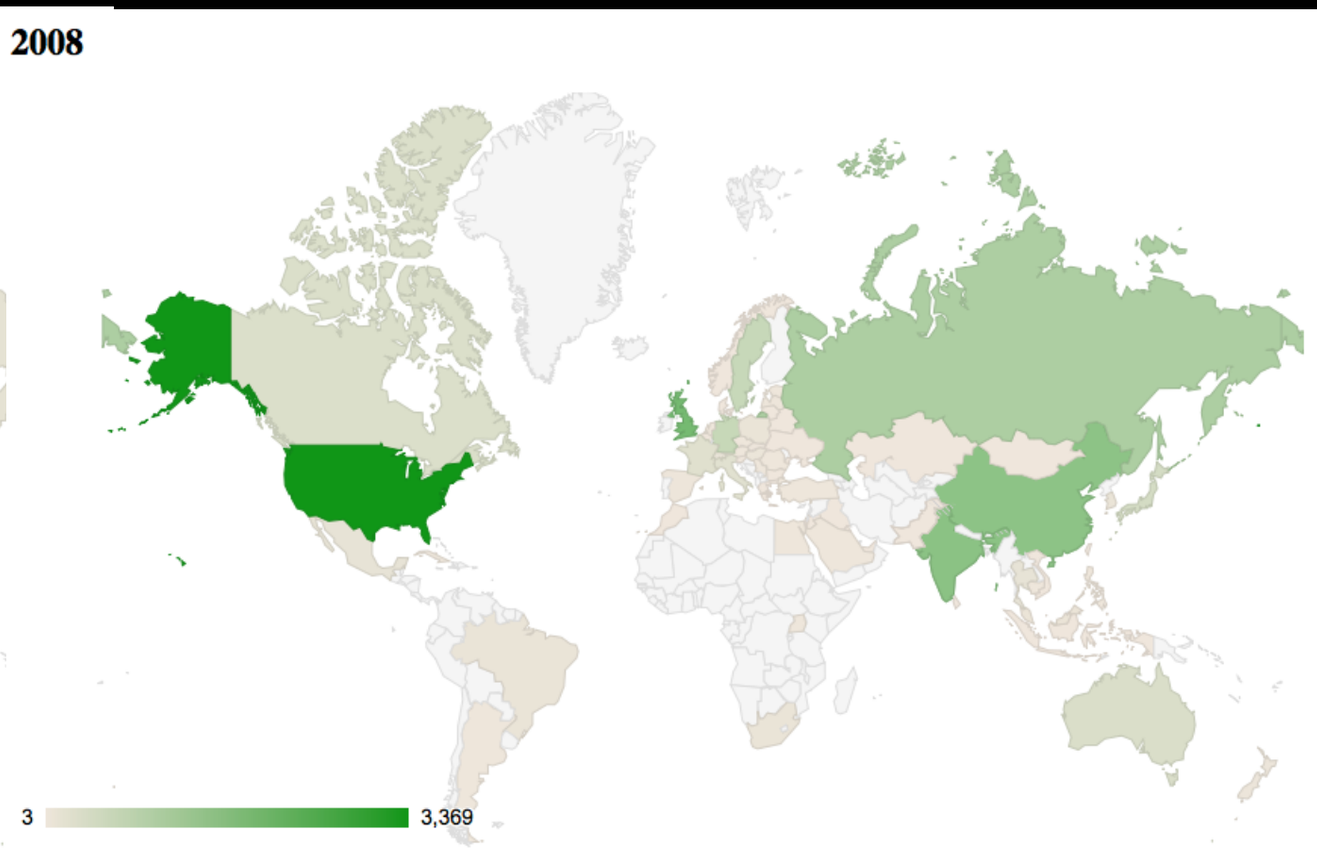
- Virtual machines with 15 modules for English and Spanish:
  - http://ixa2.si.ehu.es/nrdemo/demo.php
  - http://ixa2.si.ehu.es/nrdemo_es/demo.php
  - http://ic.vupr.nl/~ruben/vua-eventcoreference.ttl/
- Modules for Dutch and Italian
- KnowledgeStore and populators: https://www.youtube.com/watch?v=if1PRwSIl5c
- End-user interfaces to deal with large complex graphs

Semantic Web filtering



CARS: WHERE & WHEN





Provenance statistics

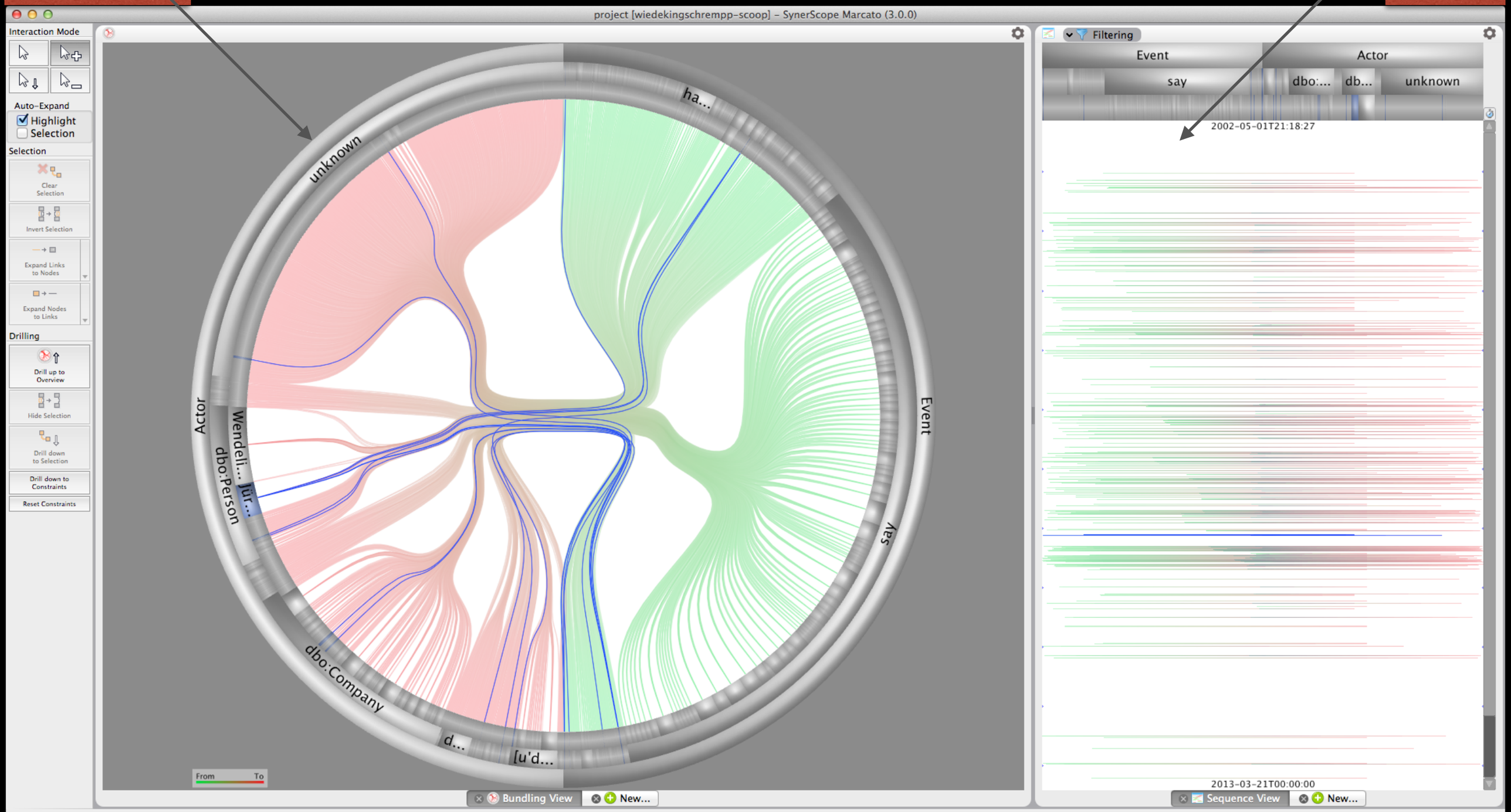
Source owner	Triples
Automotive_News	321,321
PR_Newswire	201,399
Detroit_Free_Press_(Michigan)	193,420
Just_-_Auto	167,735
Automotive_News_Europe	162,424
The_Associated_Press	160,911
just-auto_global_news	158,493
Associated_Press_Financial_Wire	151,971
The_Detroit_News_(Michigan)	150,383
The_Associated_Press_State_&_Local	129,248
etc.	…
TOTAL	12,851,504

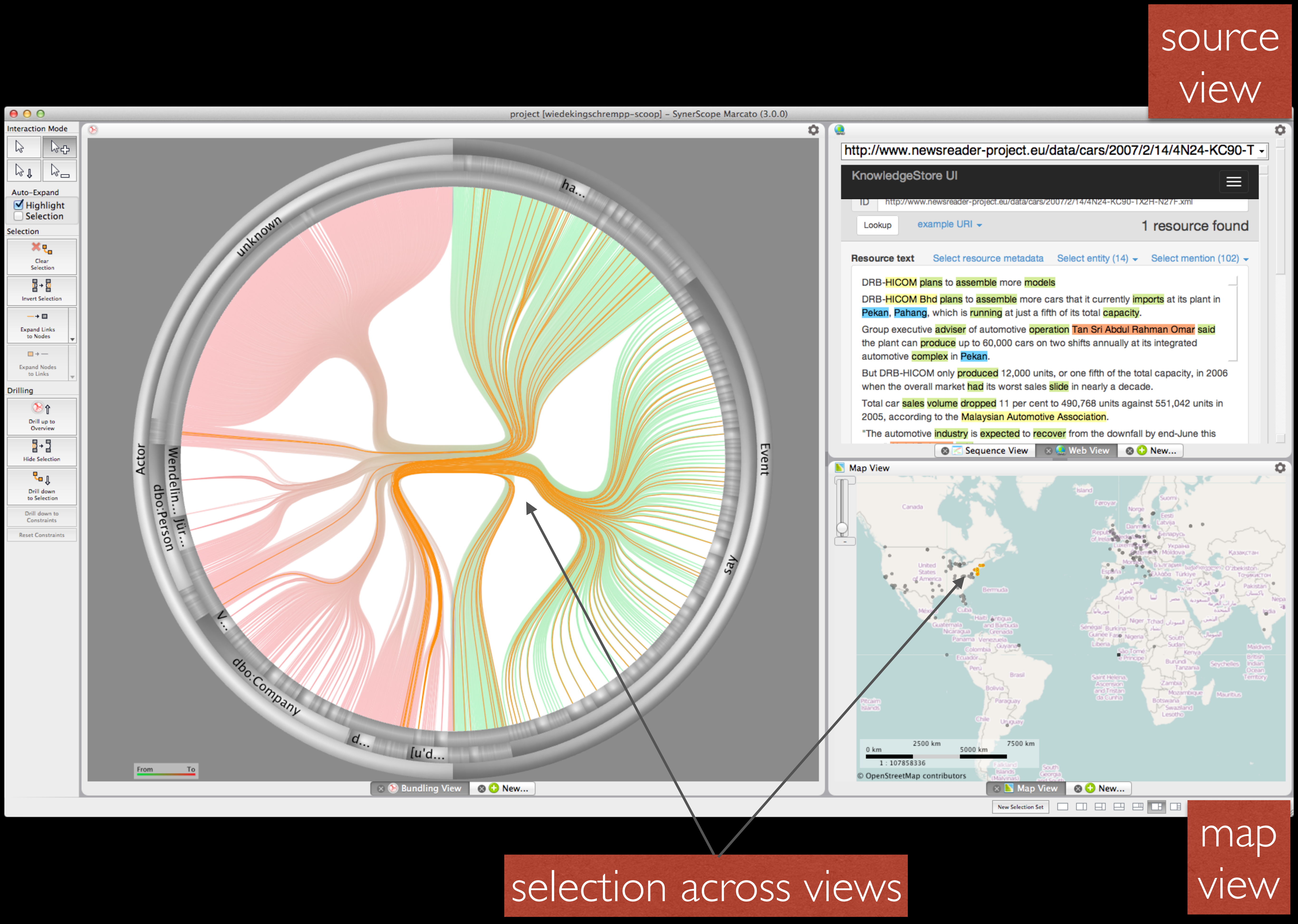
Synerscope

visualization & interaction on large knowledge graphs

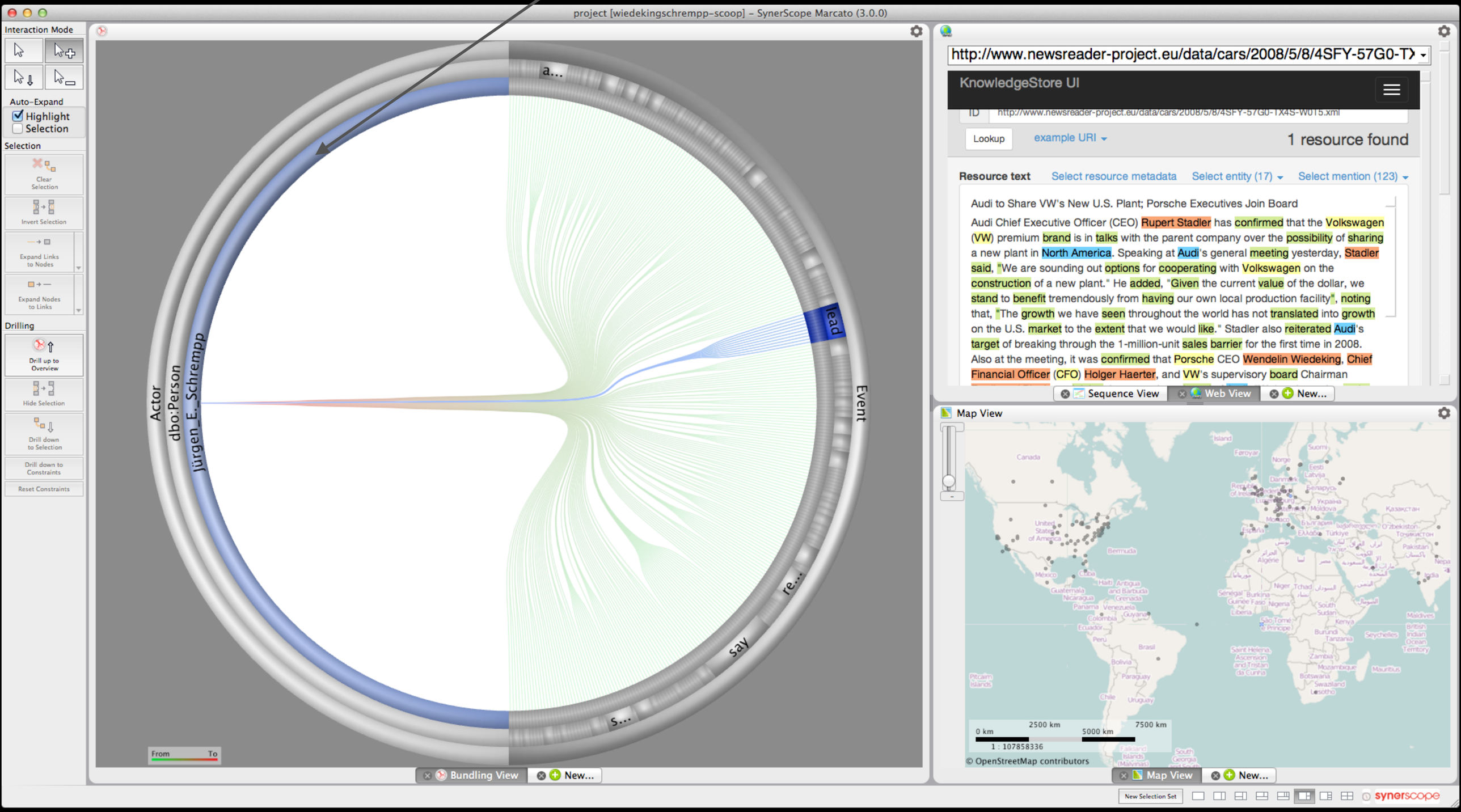
relation graph

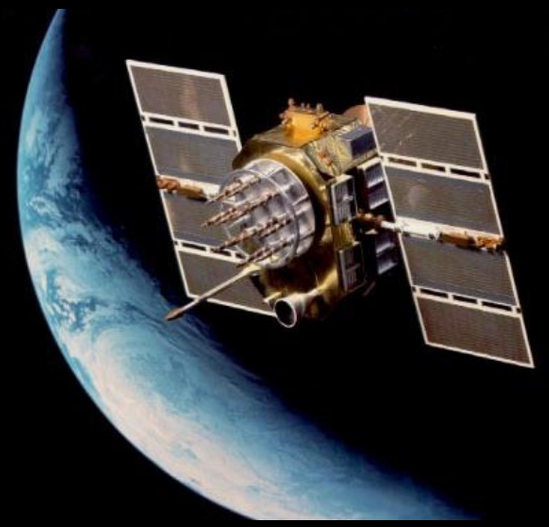
time line




source view
project [wiedekingschrempp-scoop] – SynerScope Marcato (3.0.0)
Interaction Mode
Auto-Expand
Highlight
Selection
Selection
Clear Selection
Invert Selection
Expand Links to Nodes
Expand Nodes to Links
Drilling
Drill up to Overview
Hide Selection
Drill down to Selection
Drill down to Constraints
Reset Constraints
unknown
ha...
Event
say
Actor
Wendelin... Jür...
dbo:Person
V...
dbo:Company
d...
[u'd...
From
To
Bundling View
New...
http://www.newsreader-project.eu/data/cars/2007/2/14/4N24-KC90-T
KnowledgeStore UI
Lookup
example URI
1 resource found
Resource text
Select resource metadata
Select entity (14)
Select mention (102)
DRB-HICOM plans to assemble more models
DRB-HICOM Bhd plans to assemble more cars that it currently imports at its plant in Pekan, Pahang, which is running at just a fifth of its total capacity.
Group executive adviser of automotive operation Tan Sri Abdul Rahman Omar said the plant can produce up to 60,000 cars on two shifts annually at its integrated automotive complex in Pekan.
But DRB-HICOM only produced 12,000 units, or one fifth of the total capacity, in 2006 when the overall market had its worst sales slide in nearly a decade.
Total car sales volume dropped 11 per cent to 490,768 units against 551,042 units in 2005, according to the Malaysian Automotive Association.
"The automotive industry is expected to recover from the downfall by end-June this
Sequence View
Web View
New...
Map View
0 km
2500 km
5000 km
7500 km
1 : 107858336
© OpenStreetMap contributors
Map View
New...
New Selection Set
selection across views
map view

drill down to a single individual





Demonstration event

Protests in Cairo's Tahrir Square were brought by citizen journalism. Photograph: Mohammed Abed/AFP/Getty Images