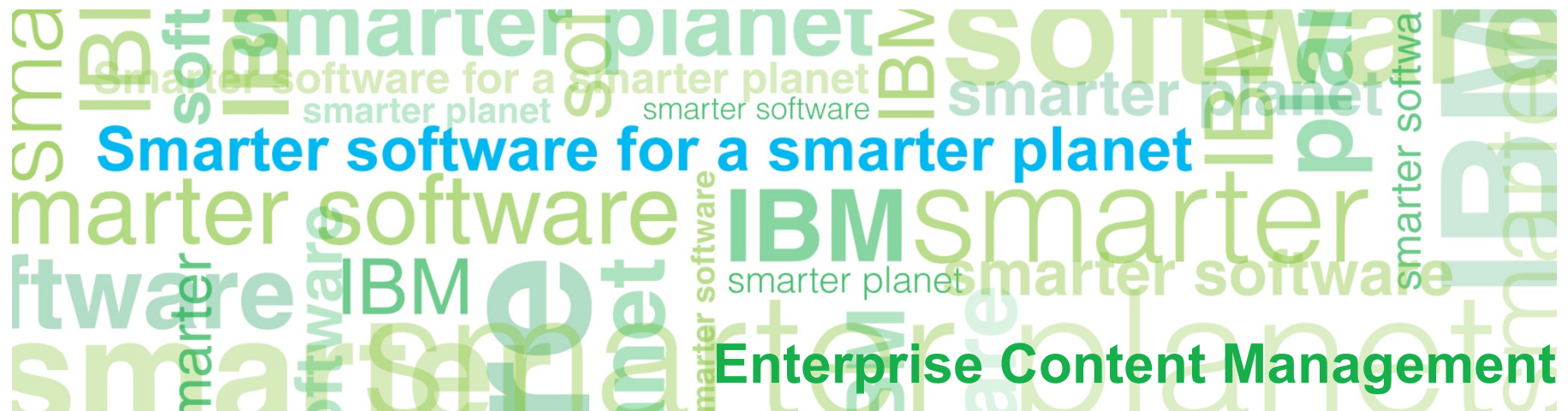


# IBM Enterprise Content Management Solutions

## *Content Analytics*



## Mi a tartalom menedzsment ?

- **Strukturálatlan adatok közötti összefüggések feltárásával foglalkozik**
  
- **Példa a strukturálatlan adatokra:**
  - E-mailek
  - Hibajegyek különböző ticket rendszerekben
  - Panasz bejelentések ( e-mail, vagy dokumentum formátiumban)
  - Műszaki dokumentációk
  - Cset, twitter, blog stb... (social media)
  - Feldolgozott hanganyag...

# Nyelvbányász projekt

SZTAKI: Projekt információk - Mozilla Firefox

Fájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

http://www.sztaki.hu/kereses/projektek/projekt\_informaciok/?uid=00104

Google

Legtöbbször látogatott Bevezetés Friss hírek SZTAKI Szótár: szotar.s...

nyelvbányász - Google keresés x NKFP Nyelvbányász projekt x SZTAKI: Projekt információk x

## KERESÉS

- Honlapon
- Telefonkönyv
- Témakörök szerint
- Projektek
- Publikációk

## OLDALTÉRKÉP

### Projekt vezető

**Szepesvári Csaba**

**Cím:** 1111 Budapest, Kende u. 13-17.  
**Szoba:** K 303  
**Telefon:** +36 1 279 6262  
**E-mail:** szcsaba@sztaki.hu  
**Honlap:** <http://www.sztaki.hu/~szcsaba>

[További információk]

**Benczúr András**



**Cím:** 1111 Budapest, Lágymányosi u. 11.  
**Szoba:** L 412  
**Telefon:** +36 1 279 6172, +36 1 279 6290  
**Fax:** +36 1 209 5269  
**E-mail:** benczur@ilab.sztaki.hu  
**Honlap:** <http://www.ilab.sztaki.hu/~benczur/>

[További információk]

### Résztvevők

MTA SZTAKI (Gépi tanulás és Adatbányászat és Webkeresés csoportok), ELTE (Számítógéptudományi Tanszék, Komplex Rendszerek Fizikája Tanszék), BME (Sztochasztikus Analízis Tanszék, Matematikai Intézet), MTA Nyelvtudományi Intézet, MTA SZFKI, Omega Consulting Kft., **Pont Rendszerház Rt.**

### Tevékenység

A szöveges e-dokumentumok elérhetővé tétele kritikus eleme a vállalatok hatékony működtetésének. A NYELVBÁNYÁSZ projekt célja, hogy megcélolja a világszinten jelentkező ilyen irányú igények kielégítését egy a ma szokásostól radikálisan eltérő új megközelítésre építve, az önszervező tanulás segítségével felépített nyelvi rendszerek alkalmazásával. Szemben a korábbi megközelítésekkel, az önszervező módon tanult nyelvmodellek jóval kevesebb ad-hoc elemet tartalmaznak, s így általuk a korábbiaknál jobb eredmények érhetőek el. A projekt keretén belül a számítógépes nyelvészek, matematikusok, kognitív tudósok, fizikusok, adatbányászok, gépi tanulás szakemberek dolgoznak együtt - a multidiszciplináris megközelítéstől is várjuk a projekt újszerű, áttörést jelentő eredményeit. Szintén cél a tanult nyelvmodellekre szervesen ráépülő ipari igényeket kiszolgáló nyelvtechnológiák, illetve az ezekre épülő, tipikus vállalati problémákat megcélzó alkalmazás prototípusok kifejlesztése is. A munkába bevont ipari partnerek és végfelhasználók közreműködése biztosítja, hogy a hasznosításkor megjelenő felhasználói igények a projekt megvalósításának kezdetétől fogva reprezentálva legyenek.

Kész

SMARTER PLANET SOTI

IBM

NKFP Nyelvbányász projekt - Mozilla Firefox

Fájl Szerkesztés Nézet Előzmények Könyvjelzők Eszközök Súgó

http://nyelvbanyasz.sztaki.hu/

Google

Legtöbbször látogatott Bevezetés Friss hírek SZTAKI Szótár: szotar.s...

nyelvbányász - Google keresés NKFP Nyelvbányász projekt SZTAKI: Projekt információk

# NKFP 2004, Nyelvbányász projekt

## Résztvevők

1. MTA SZTAKI (Koordinátor); [Gépi tanulás csoport](#) és [Adatbányászat és webkeresés csoport](#)
2. ELTE ([Számítógéptudományi Tanszék](#), Komplex rendszerek fizikája [Pollner Péter](#), IK [Istenes Péter](#))
3. BME Mat. Intézet, [Sztochasztika Tanszék](#)
4. [MTA Nyelvtudományi Intézete](#)
5. MTA SZFKI, [Fáth Gábor](#)
6. [Omega Consulting](#)
7. [Pont Rendszerház](#)

## Szemináriumok - 2005 januárig

1. Dec. 17  
Pollner Péter előadása: [Alignment Based Learning](#), Menno van Zaanen nyomán
2. Dec. 3.  
[Bíró István előadása](#)
3. Nov. 26; Goodman, Charniak: The State of the Art in Language Modeling. I.rész; Tutorial Presented at AAAI 2002 [ppt](#)  
A Bit of Progress in Language Modeling, Extended Version Microsoft Research Technical Report MSR-TR-2001-72 [pdf](#)
4. Nov. 12  
Balog Krisztián: [klasszifikáció](#),  
Schönhofen Péter: [feature selection](#)  
Szepesvári Csaba: [további tematikára javaslat](#)
5. Okt. 29  
Szepesvári Csaba: projektmegbeszélés; [ppt](#)
6. Október 21: Kálmánnal egyeztetés NYTI feladatokról, memó [itt](#)
7. Okt. 1.  
Fóris Zoltán: reprezentáció és tanulás; [előadás](#) és [doc file](#).
8. Szeptember eleje

Kész

A mai információs világban a tudás felhasználása üzleti előnyt jelent ...



**Gépesített**



**Behálózva**



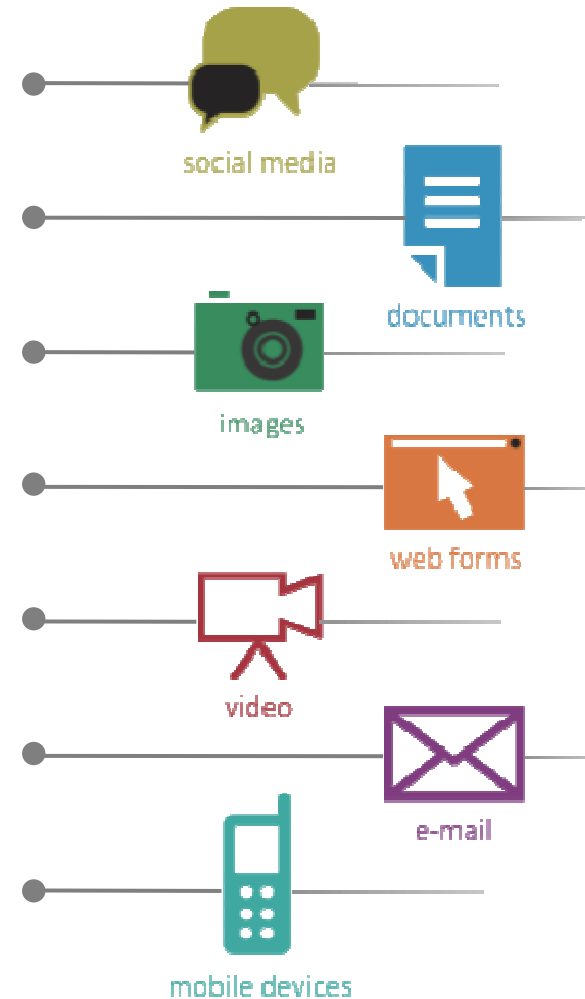
**Intelligensebb**

*... ehhez szükséges a jobb  
Enterprise Content Management  
Vállalati tartalom menedzsment*

- Kezeljük az alapvető információkat bárhol is vannak
- Irányítsuk a információ áramlását a teljes élelciklus alatt
- Optimalizáljuk a tartalommal kapcsolatos folyamatokat
- Találjunk rá a váratlan összefüggésekre

# Találjuk meg a **jelet** a zajban

A tartalom hasznosításához szükséges hogy tudjunk benne **keresni**, tudjunk **elemezni** és **értékelni** nagy mennyiségű szöveget dokumentumot, hogy megértsük és felismerjük a lényeges információt bármilyen forrásból jön is...



# IBM Content Analytics adds value to ...



## Healthcare Analytics

- **Analyzing:** E-Medical records, hospital reports
- **For:** Clinical analysis; treatment protocol optimization
- **Benefits:** Better management of chronic diseases; optimized drug formularies; improved patient outcomes



## Customer Care

- **Analyzing:** Call center logs, emails, online media
- **For:** Buyer Behavior, Churn prediction
- **Benefits:** Improve Customer satisfaction and retention, marketing campaigns, find new revenue opportunities



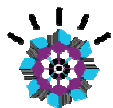
## Crime Analytics

- **Analyzing:** Case files, police records, 911 calls...
- **For:** Rapid crime solving & crime trend analysis
- **Benefits:** Safer communities & optimized force deployment



## Insurance Fraud

- **Analyzing:** Insurance claims
- **For:** Detecting Fraudulent activity & patterns
- **Benefits:** Reduced losses, faster detection, more efficient claims processes



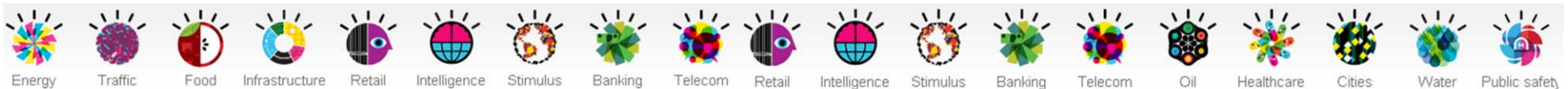
## Automotive Quality Insight

- **Analyzing:** Tech notes, call logs, online media
- **For:** Warranty Analysis, Quality Assurance
- **Benefits:** Reduce warranty costs, improve customer satisfaction, marketing campaigns



## Social Media for Marketing

- **Analyzing:** Call center notes, SharePoint, multiple content repositories
- **For:** churn prediction, product/brand quality
- **Benefits:** Improve consumer satisfaction, marketing campaigns, find new revenue opportunities or product/brand quality issues



US Government Agency

# Smart is: **intelligently classifying** documents

*“Consistent, reliable and automated configuration content is critical.”*



*Industry context: government*

*Value driver: speed, accuracy of classification*

*Solution onramp: content analytics*

### **Business Challenge**

With millions of email messages going through this government agency’s systems every year, the agency needed to improve the accuracy and speed of its content categorization in order to meet regulations for accurate and effective records retention.

### **What’s Smart?**

The agency transformed its manual, inaccurate human categorization process with automated classification technology. The agency resolved inconsistencies in content categorization using IBM Classification Module’s contextual classification, replacing its over-burdened, labor-intensive content categorization process.

### **Smarter Business Outcomes**

Improves visibility and access to accurately classified email content. Provides more insight for records retention and legal discovery. Reduces storage required for email messages.



## Japan Business Services Provider

# Smart is: **gleaning insight** about customers

*“Insight into customer interaction logs is an information gold mine for us.”*

— General Manager  
Japan



*Industry context: computer services*  
*Value driver: improve customer service*  
*Solution onramp: content analytics*

### **Business Challenge**

A Japanese business services provider operates multiple customer service centers and needed ways to analyze large volumes of information to improve agent training and deliver better customer support.

### **What’s Smart?**

They implemented content analytics from IBM to understand and process natural language. The solution analyzes customer interactions based on consolidated logs of phone calls, email and Web, identifying keywords.

### **Smarter Business Outcomes**

Improved agent skills and training, resulting in a 92% reduction in call transfer and 88% improvement in volume. Provides new insights about product issues, resulting in an 88% decrease in product-related calls.

Financial institution

# Smart is: creating rapid insights from content

*Industry context: banking and financial services*  
*Value driver: internet fraud prevention*  
*Solution onramp: content analytics*



*“The demo impressed the customer so much that the customer was ready to buy ICA in a few days,” ECM Sales Rep.*

### Business Challenge

A European financial Institution wanted to investigate fraudulent behavior by exploring internet sites for actions that might pose a threat to its members.

### What’s Smart?

In less than one week, using IBM Content Analytics, the IBM sales team analyzed a selected set of websites, investigated their findings and reported their findings back to the customer.

### Smarter Business Outcomes

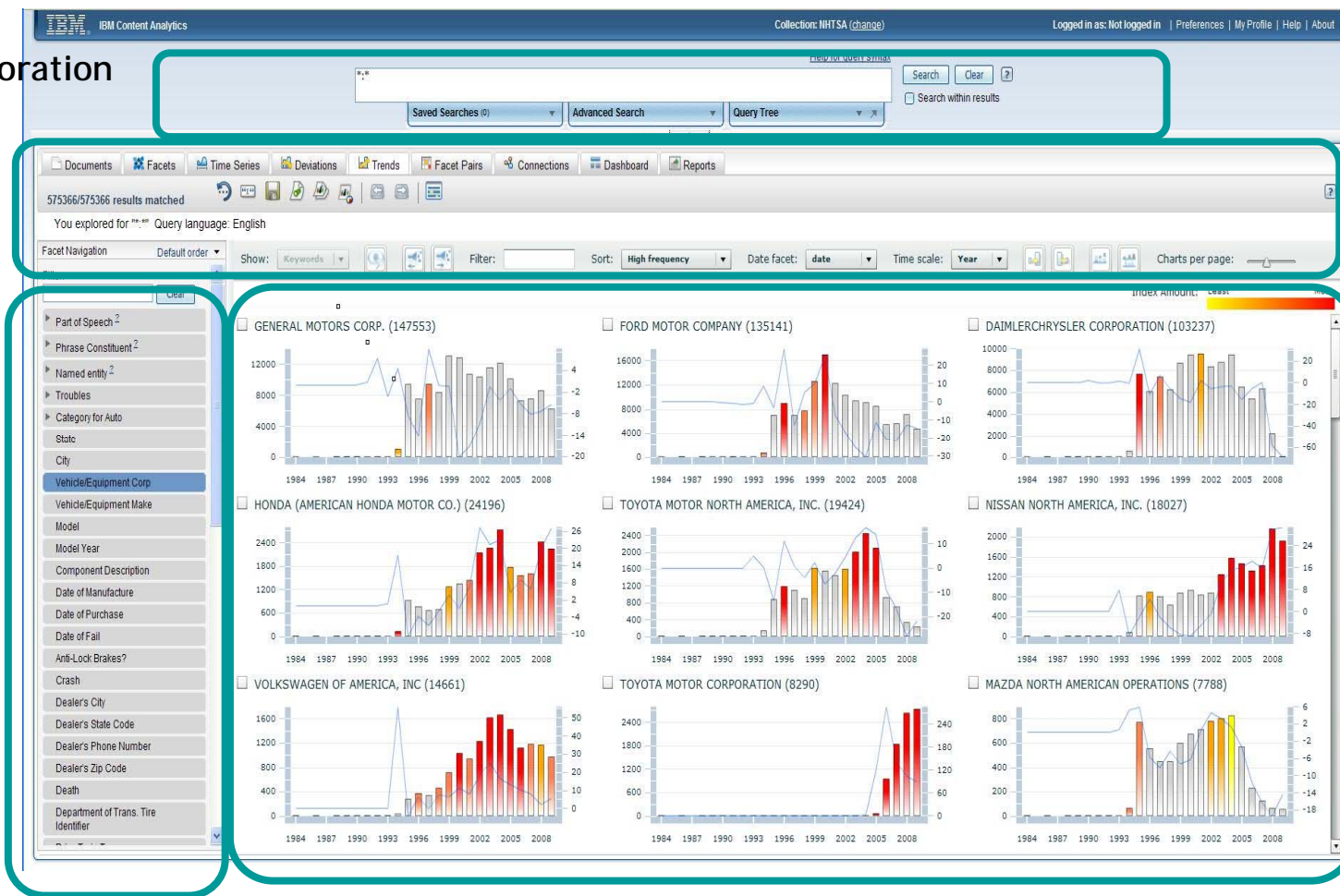
The team rapidly showed the customer types of intrusion correlating bank terms with news about a known hacker using the out of the box extraction capabilities, prevention scenarios and frequently vulnerable operation systems.

## The Interactive Discovery User Interface Explained

Search Query Exploration

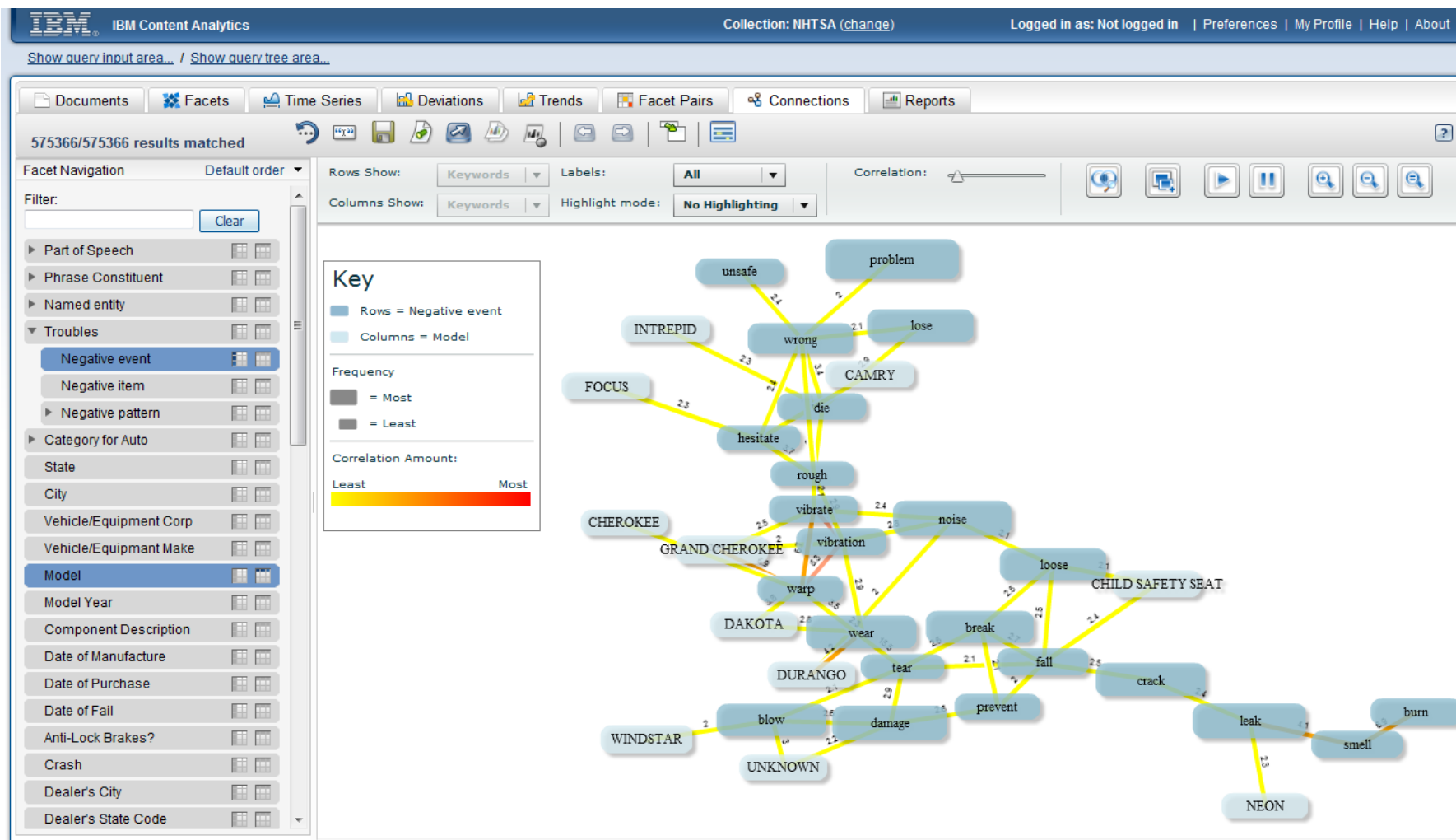
Views, Filters and Thresholds

Automatically  
Extracted and  
Analyzed  
Concepts, Entities,  
Relationships,  
Meta Data and  
Classifications

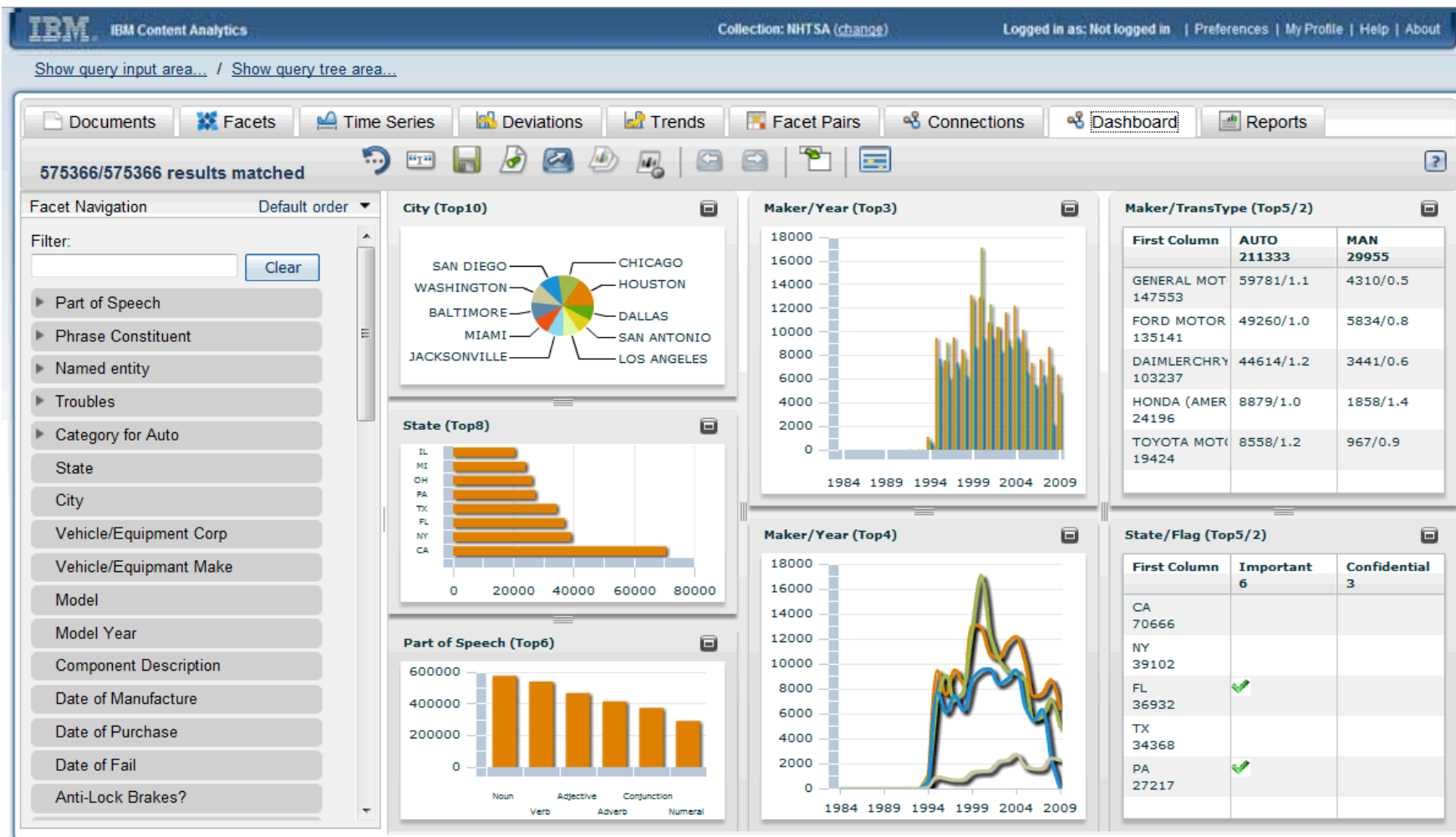


Visualization with Drill Down for Exploration and Assessment

## Connections View links highly correlated terms to one another



## Create Dashboard Views for Executive Summaries



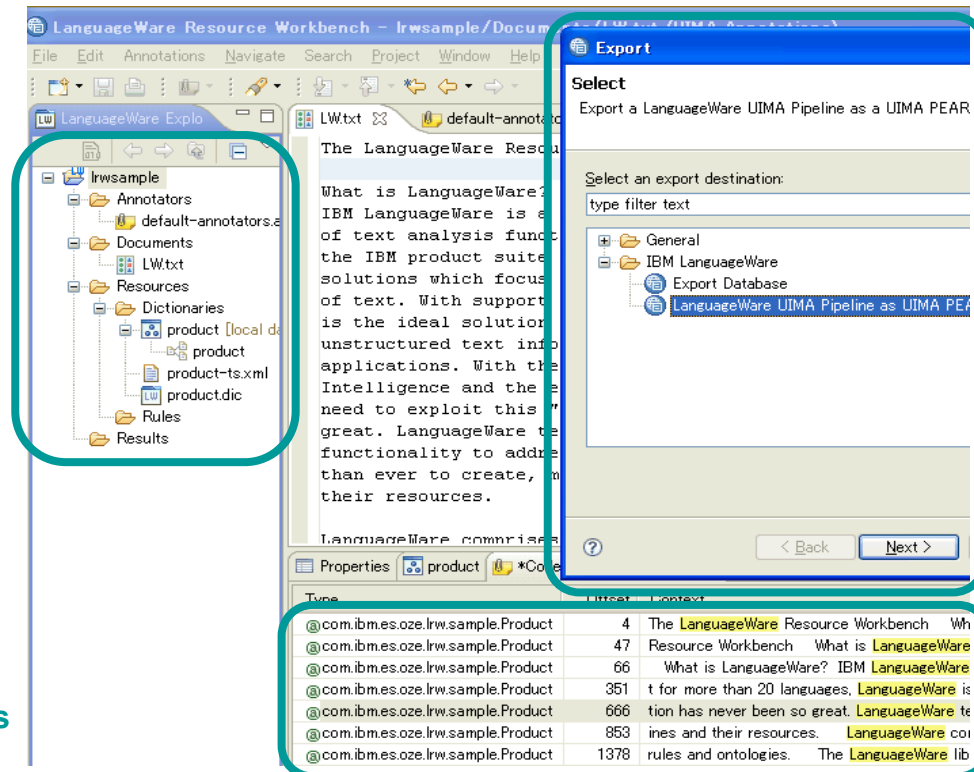
# Steps to tailor your text analysis with flexible, easy-to-use tooling

## 1 Develop your Custom Text Analysis with Tooling

Build language and domain resources into a LanguageWare dictionary.  
 Develop rules to spot facts, entities and relationships.  
 Create and test UIMA annotators with a collection of documents.

## 2 Export your Custom Text Analysis

Easily generate the annotators to be Content Analytics ready



View of Project Resources

Easy to export your custom text analysis

Easy to test and verify your tailored text analysis

## 3 Deploy your Custom Text Analysis with in CCA

Import newly created annotators via Content Analytics administration console and associate it to a collection.