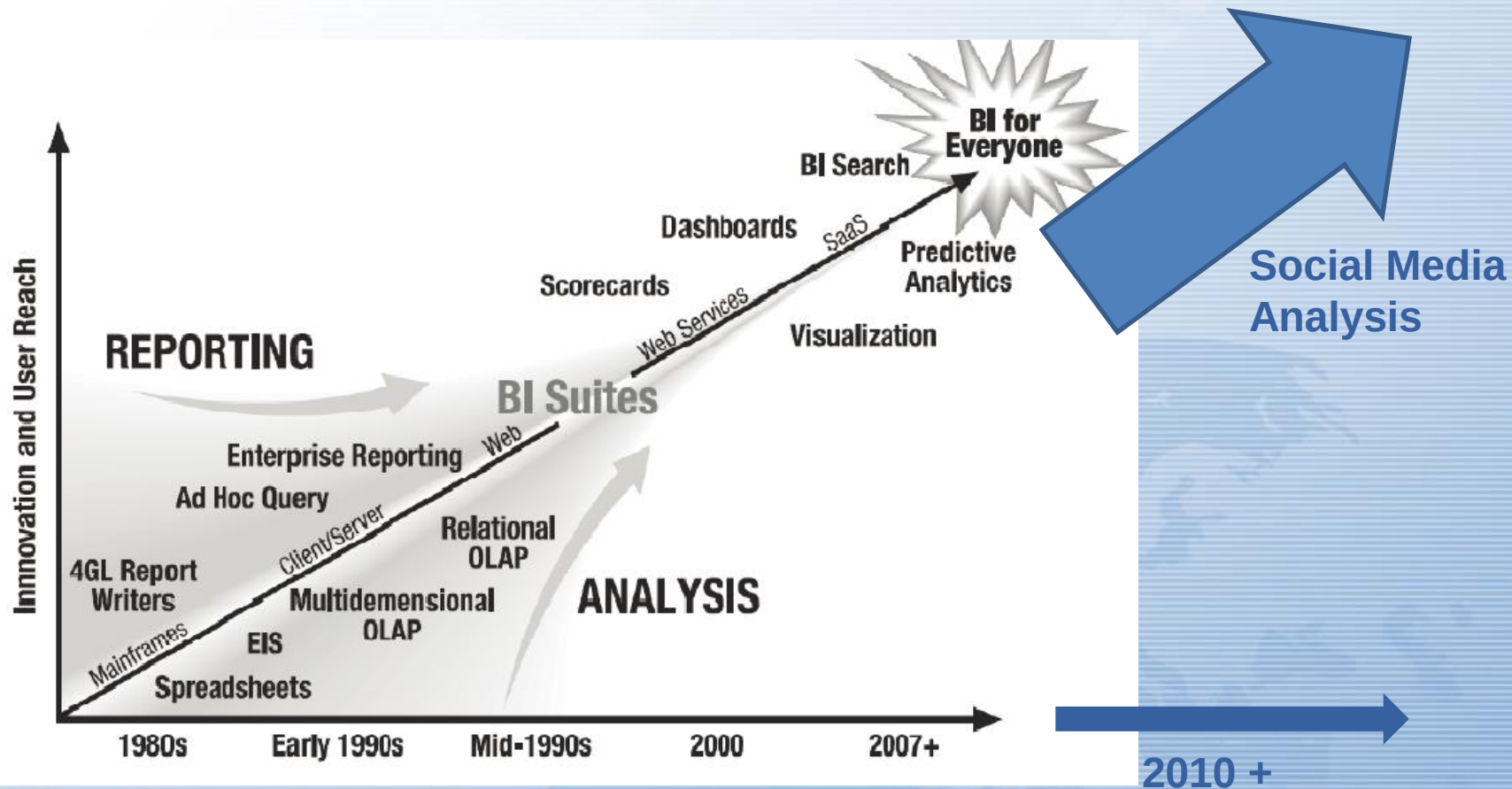
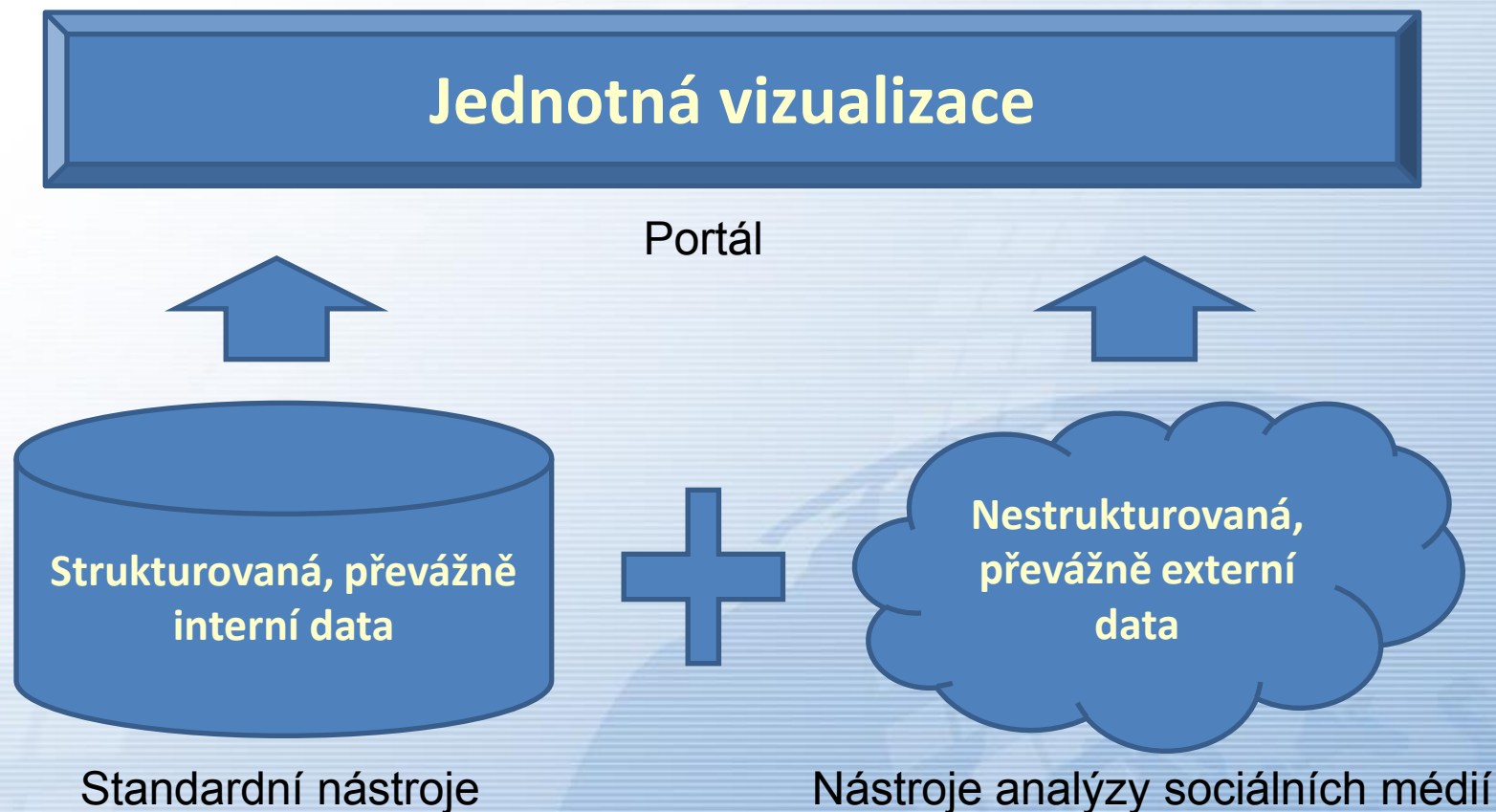
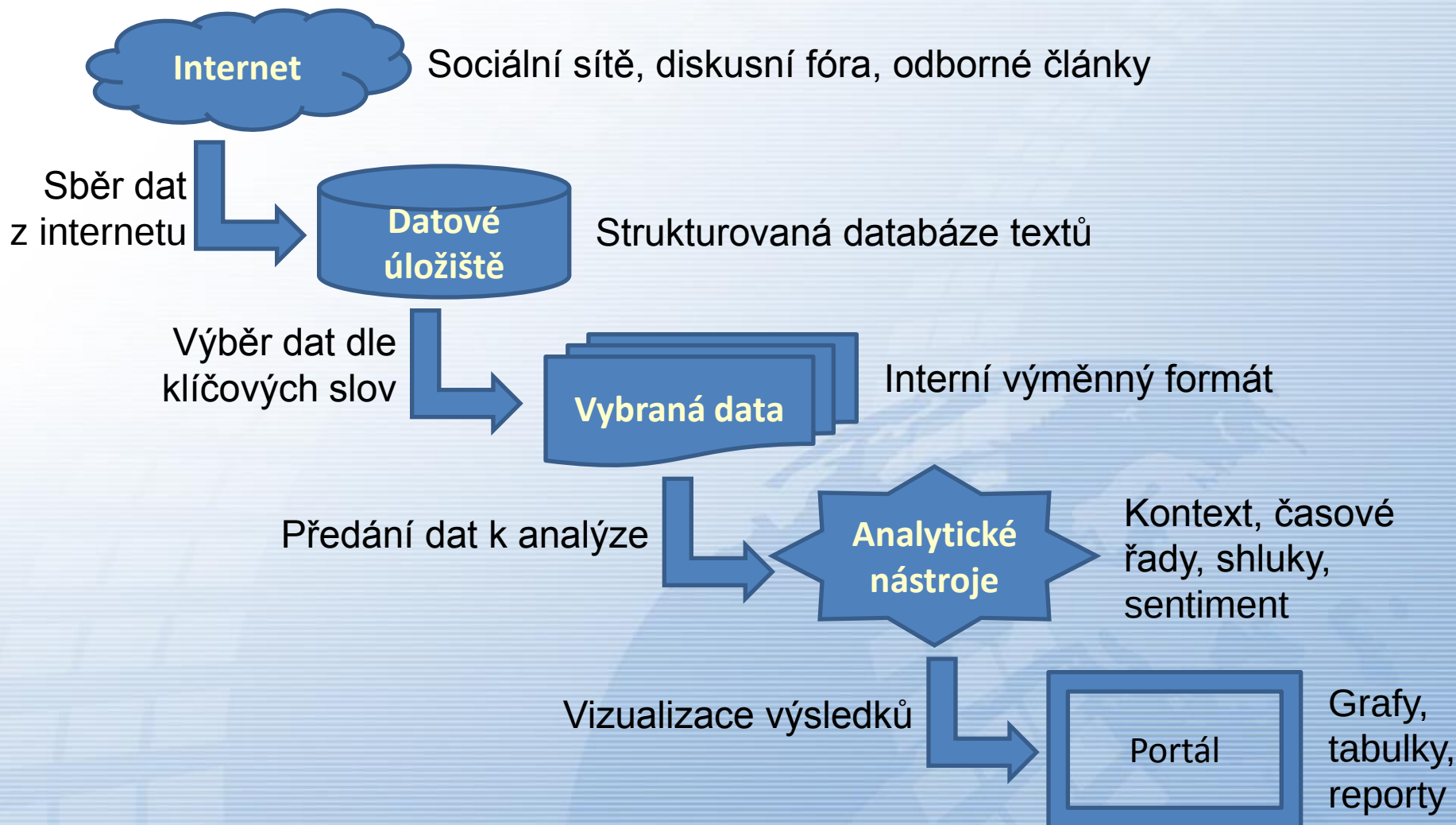


TECHNIKY ZÍSKÁVÁNÍ DAT Z INTERNETU PRO POTŘEBY POLICIE

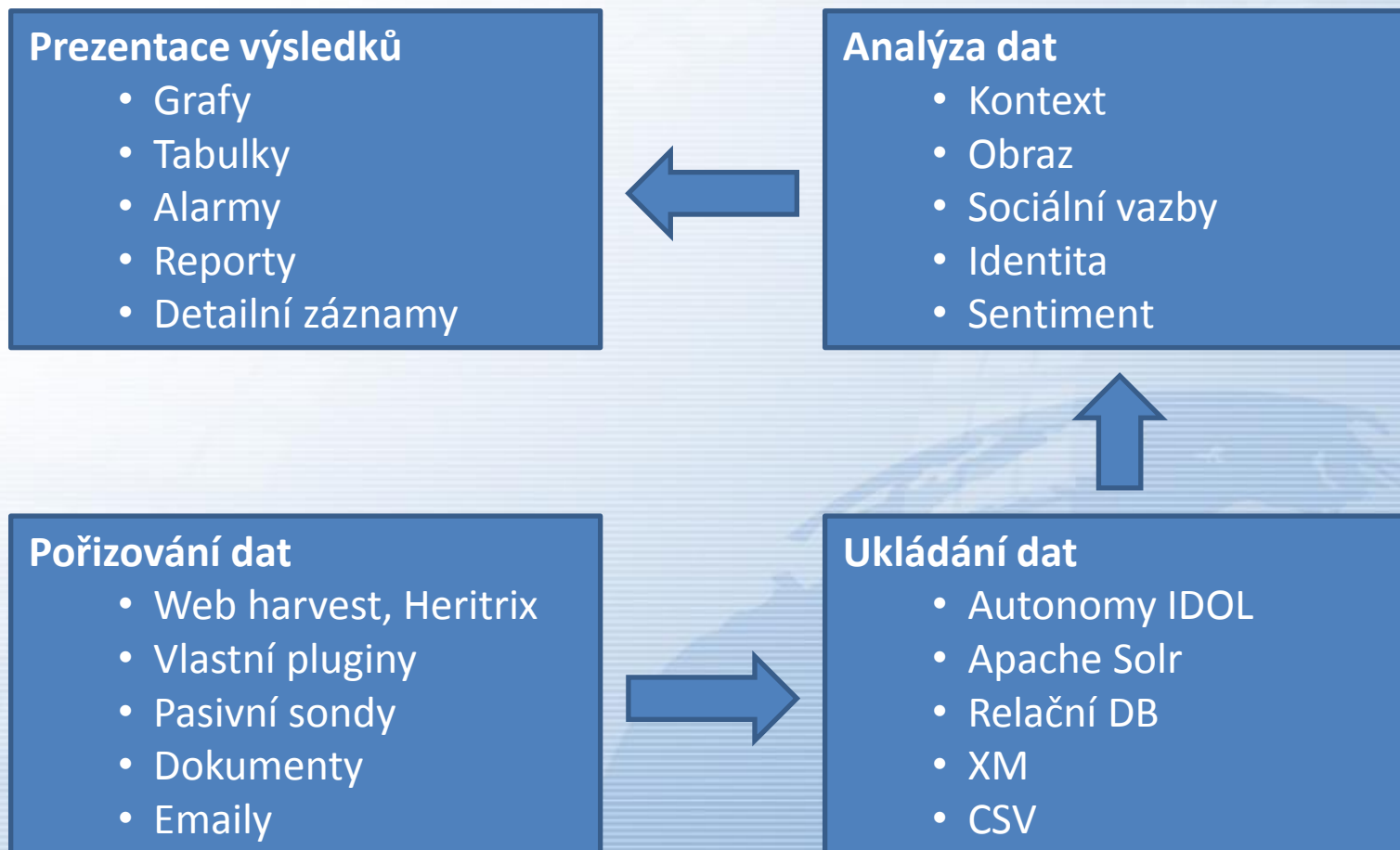
Jaroslav Ráček, Dalibor Toth



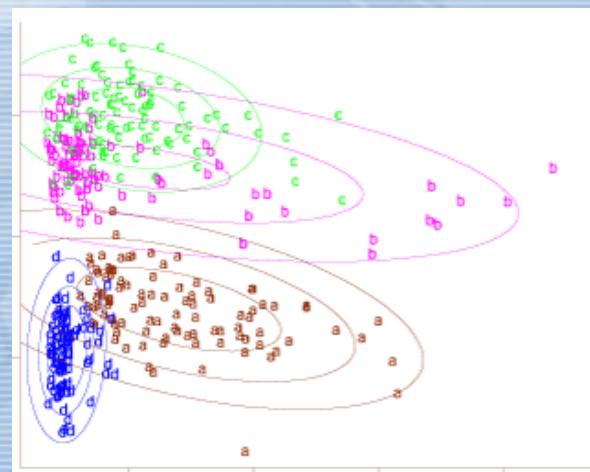
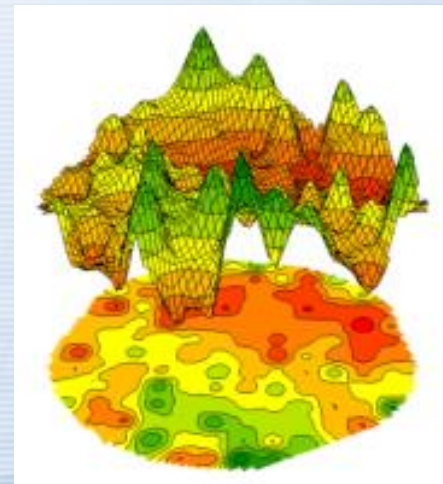




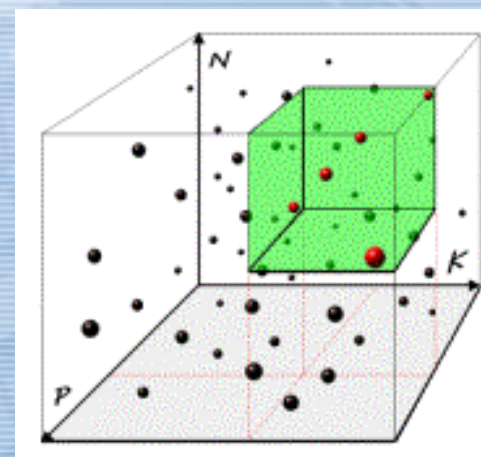
- Rozpoznat hledanou entitu, tj. předmět, osobu nebo událost,
- Rozpoznat vztahy mezi nalezenými entitami,
- Rozpoznat vztah nalezené entity nebo skupiny entit k okolnímu světu.



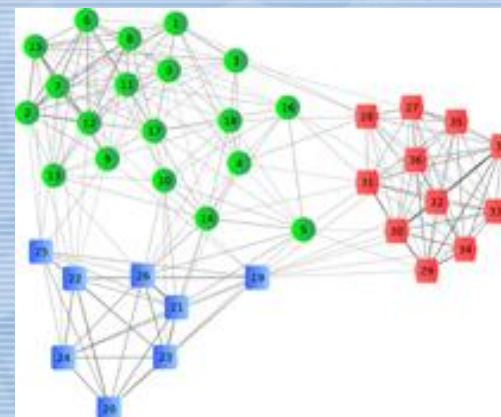
- Cílem je identifikovat klíčové výrazy, o kterých se v textu mluví.
- V úvodním předzpracování rozpoznat jazyk a opravit překlapy a chyby.
- Berou se v úvahu pouze slova nesoucí význam.
- Shlukování podle synonym.
- Hledání entit v textu na základě referenčního vzorku.



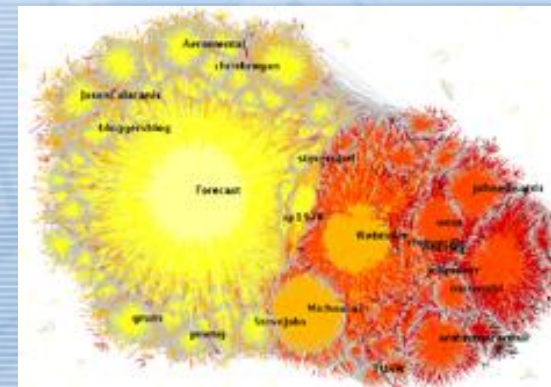
- Jako vstup pro tyto analýzy neslouží pouze obraz hledané entity, ale vstupuje i sémantická textová informace.
- Algoritmy rozpoznávání obrazu pak díky této informaci mohou scénu lépe rozdělit a hledaný předmět rozpoznat s větší přesností.
- Vstupní sémantické informace mají různou strukturu v závislosti na prohledávané předmětné doméně a zdroji dat.
- Používáme externí knihovny našich partnerů.
- Výstupem je vektor relevance.



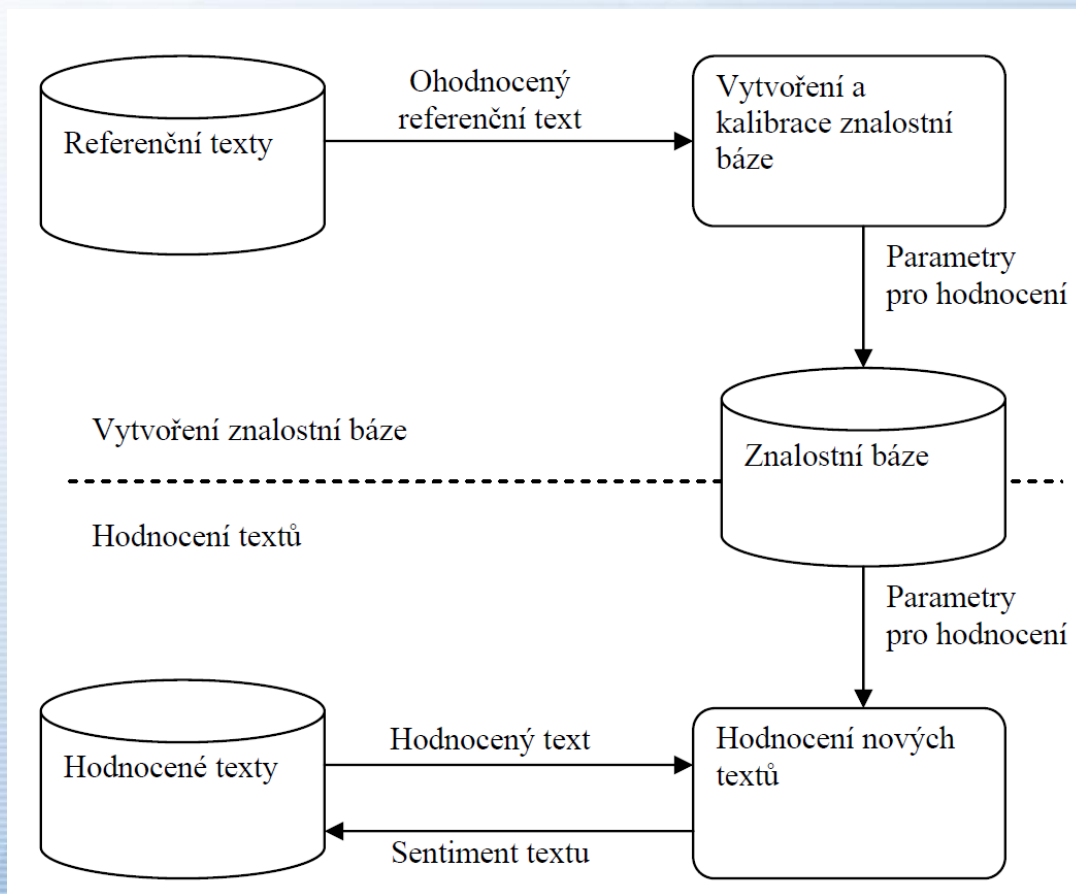
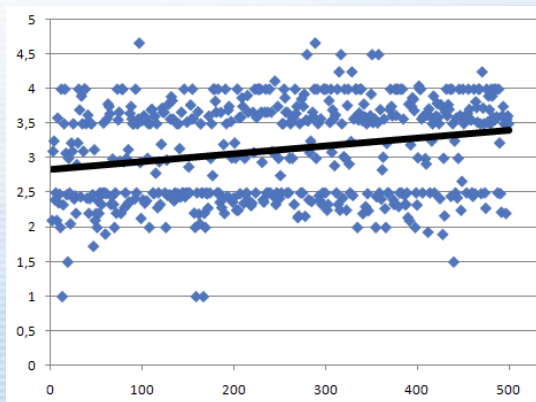
- V datech jsou rozpoznávány osoby.
 - Autoři nebo osoby, o kterých se mluví.
- Ze záznamů o osobách z různých zdrojů, je rekonstruována interní „sociální“.
- U každé osoby jsou evidována témata a další osoby se kterými je ve vztahu.
- Jsou rozlišovány typy vztahů a jejich intenzita.
- Jako samostatný parametr je evidován čas.
- Provádí se segmentace do skupin, k dispozici jsou metriky určující blízkost jednotlivých osob



- Osoby používají celou řadu pseudonymů nebo vystupují často zcela anonymně.
- Za více identitami se často skrývá stejná osoba.
- Lze částečně rozpoznat ze strukturovaných dat.
 - Shodné kontaktní údaje.
 - Většinou ale nejsou k dispozici.
- Je sledována intenzita, doba a frekvence aktivit jednotlivých pseudonymů.
- Probíhá analýza napsaných textů.
 - Podobnost slovní zásoby,
 - Častý výskyt vybraných slovních kombinací,
 - Shodné pravopisné chyby,
 - Podobné překlady.



Základním principem analýzy je hledání specifických slovních kombinací, jež indikují příslušný typ a míru sentimentu v dané doménové oblasti.




[Sign In](#)

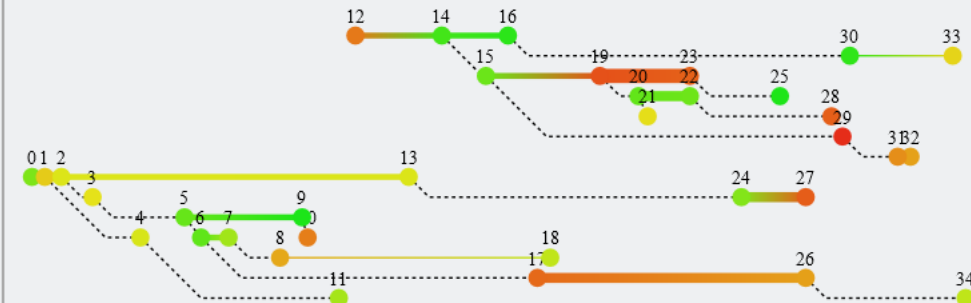
Welcome

Liferay > Welcome

Spectrogram

Jump to date:  Branch spacing: 110
 Initial view: Data source: [Reload SVG](#)

3.12.2011 9:15:29



Diagram

Selected date:

[Reload SVG](#)

Current Users



Bruno (Admin)

The admin has full control over the entire portal, allowing modification and creation of users, communities, and roles.

[View bruno's public page](#)

[Login as bruno](#)



Richard Publisher

Richard has article submission rights for the content management system.

[View richard's public page](#)

[Login as richard](#)



Michelle Editor

Michelle has control over the staging and layout of all articles as well as publishing rights for workflow and the content management system.

[View michelle's public page](#)

[Login as michelle](#)



John Regular User

John is a user with minimum rights within the portal.

[View john's public page](#)

[Login as john](#)

[Sign In](#)

Email Address

Password

☐ Remember Me

[Sign In](#)

The screenshot displays a Liferay portal interface with a search results page. The search query is "android". The results are displayed in a "Flash Visualization" chart, which shows a line graph with five data points. The y-axis represents a score, ranging from 0 to 4,470. The x-axis represents a sequence of results. The chart shows a general upward trend, peaking at the fourth data point, followed by a slight decline at the fifth point. A tooltip for the fifth data point indicates a value of 2,656 for the query "android".

Search Interface:

- Search:** Enter your query: android. Search in results.
- Display controls:** Basic, Parametric, Category. Database: telco. Min. relevance: 0. Min. Date: [empty]. Max. Date: [empty]. Search, Clear.
- Dynamic Tag Cloud:** long (645), join (150), erest (140), build prop (72), zasahovanie do systemu tento build (35), prop (35), tragické zasahovanie do systemu tento build (35), gameofu tam inak (35), preplisoval na viacerych (35), Railroad Barons (11), Papaya Railroad (11).
- Parametric Navigation:** No fields to be displayed. You must edit portlet preferences.
- Search Result:** Query: android. Display options. Showing 5 results out of 17,250. Rel: 0%, Docs: 5. Page 1 / 3,450.

Flash Visualization:

- Display controls:** Min. Score: 30%, Zoom: Show, 5 items, 1 / 1.
- Drill down:** MESIC.
- Chart:** Line graph showing scores for "android". The y-axis ranges from 0 to 4,470. The x-axis represents a sequence of results. The chart shows a general upward trend, peaking at the fourth data point, followed by a slight decline at the fifth point. A tooltip for the fifth data point indicates a value of 2,656 for the query "android".

Footer:

- Settings, Online Friends (0).
- Taskbar: Demoscreenshot - Dolphin, Vizualizace - Liferay - Mozilla Firefox, scr10.jpg (1921x1202) 9, LxScreenshot, Welcome - Liferay - Opera, Johnnie@arch:~ - Liferay, 16:09.

RNDr. Jaroslav Ráček, Ph.D.

jaroslav.racek@ibacz.eu

IBA CZ, s.r.o.
Petržálkova 2565/23
158 00 Praha 5

IBA CZ Development Center
Křenová 72
602 00 Brno

www.ibacz.eu