



Boolean Query Business Rules

Joseph Busch, Principal Analyst

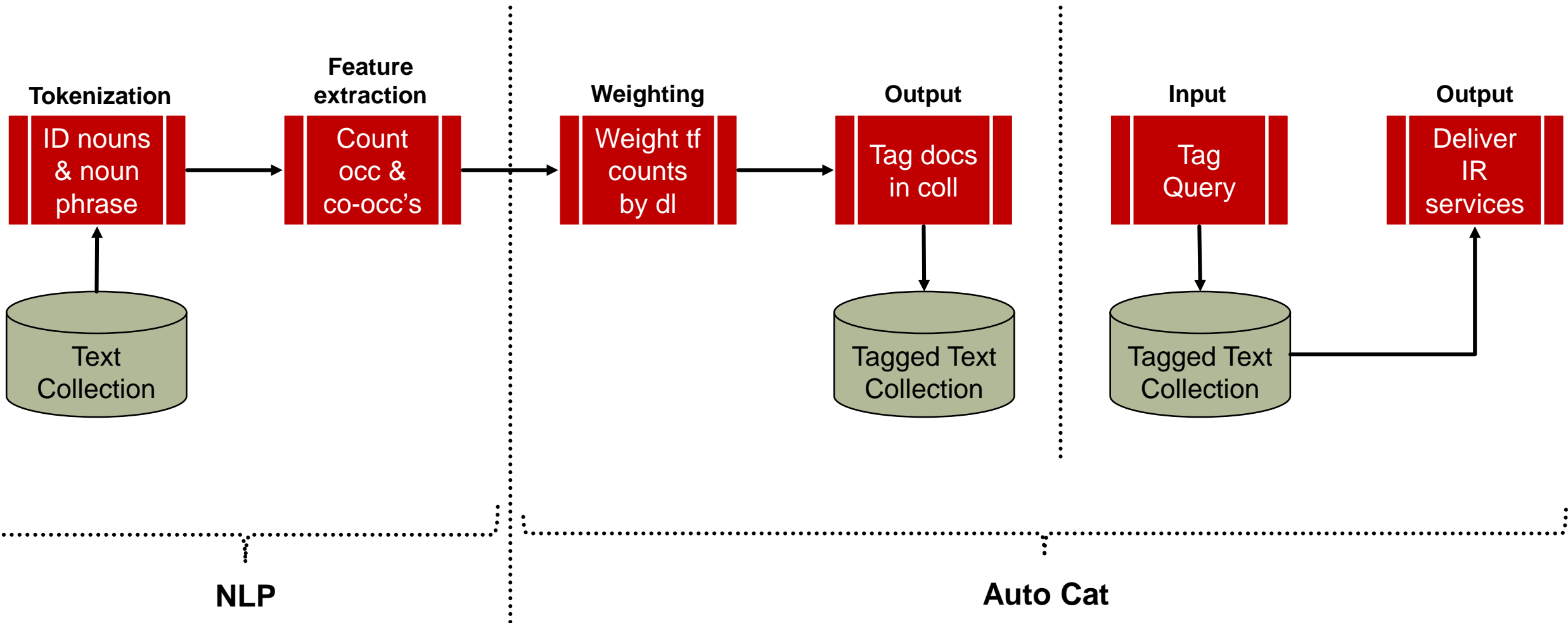


Categorization goals

- ❖ Make sense
- ❖ Provide clear perception
- ❖ Is trustworthy and as unbiased as possible



Natural language processing enables automated categorization



Automated tagging methods

- ❖ Entity extraction
- ❖ Sentiment analysis
- ❖ Keyword extraction
- ❖ Summarization
- ❖ **Predefined Boolean queries**
- ❖ Trained categorizers
- ❖ Statistical categorizers

- ❖ Some common out of the box predefined categorizers:



- ❖ But what if you want to categorize to some other scheme?

Some “Pure Play” classification tools

| Methods | Aylien | Cogito Intelligence API | IBM Watson NLP | Intellexer | Lexalytics | Meaning Cloud |
|----------------------------|--------|-------------------------|----------------|------------|------------|---------------|
| Entity extraction | X | X | X | X | X | X |
| Sentiment analysis | X | | X | X | X | X |
| Keyword extraction | X | X | X | | | |
| Summarization | X | X | | X | X | |
| Predefined Boolean queries | X | X | X | X | X | X |
| Trained categorizers | | | X | X | X | |
| Statistical categorizers | | | X | X | | |
| API | X | X | X | X | X | X |

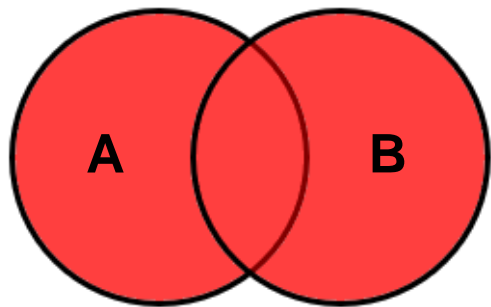
Some applications with automated classification

| Methods | Concept Searching | Data Harmony | Expert System | Mondeca | PoolParty | Smart Logic |
|----------------------------|-------------------|--------------|---------------|---------|-----------|-------------|
| Entity extraction | X | X | | | X | X |
| Sentiment analysis | | X | | | | |
| Keyword extraction | X | X | | | X | X |
| Summarization | | X | | | | |
| Predefined Boolean queries | | X | | | | X |
| Trained categorizers | | X | X | | X | X |
| Statistical categorizers | | | X | | | |
| API | | X | X | X | X | X |
| Integrations | | | | | X | X |

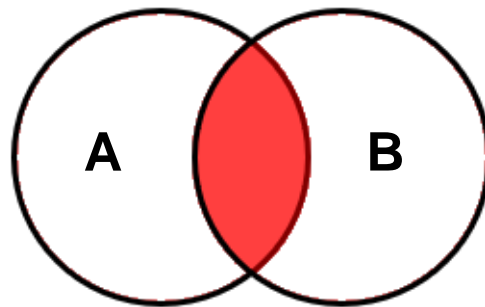
Boolean queries

- ❖ Basic operators
 - AND (conjunctive)
 - OR (disjunctive)
 - NOT (negation)

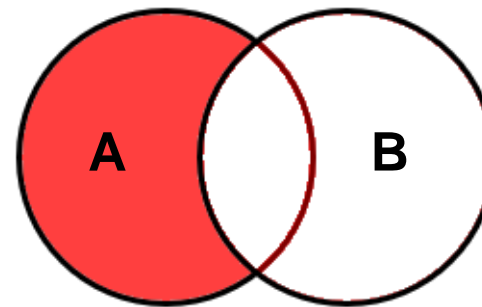
- ❖ Venn diagrams



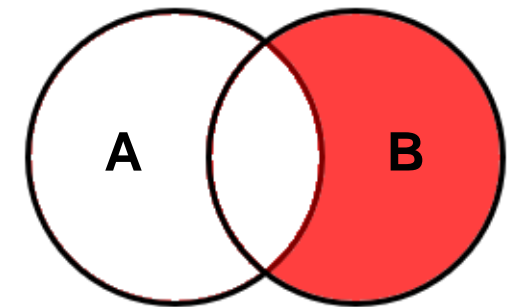
A OR B



A AND B



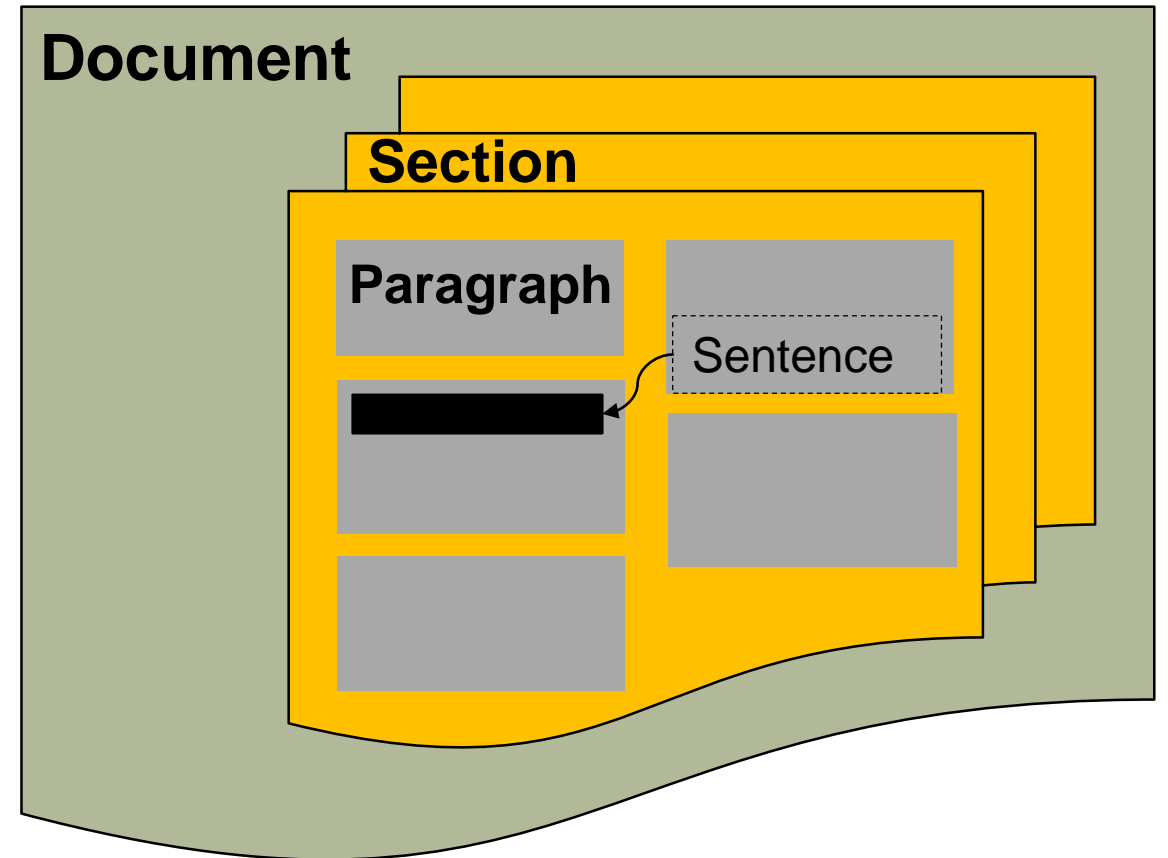
A NOT B



B NOT A

Proximity operators

- ❖ Proximity search (specified distance). Hint: Proximity operators and syntax are not standardized.
 - NEAR
 - NOT NEAR
 - FOLLOWED BY
 - NOT FOLLOWED BY
 - SENTENCE
 - FAR



Query syntax

- ❖ Bounded phrase
 - Usually quotation marks, e.g.
“health insurance”
- ❖ Truncation (right, left, internal)
 - Usually an asterisk, e.g.
child*
“pre-existing condition*”
- ❖ Nested statements
 - Parentheses (that must match up)
 (“health insurance” AND (children* OR “pre-existing condition*”))

How to create a Boolean query (1)

- 1) Brainstorm a list of 10 relevant words and phrases.
- 2) Use that list to identify 10 relevant items (articles, videos, websites, etc.)
 - E.g., do a Google search, search Google Scholar, search the NYT (or any other newspaper that you subscribe to), search Library of Congress Chronicling America (1789-1963), etc.
- 3) Review 10 relevant items and write down the words and phrases that provide a context for the theme/topic/concept.
 - Titles, headings, summaries, introductions (at the beginning) and conclusions (at the end) are good areas to focus on without having to read the whole item.
- 4) Note any named entities (people, organizations, events, laws, etc.) that are closely associated with the theme/topic/concept.
 - E.g., for gun violence “Gabrielle Giffords”, “Michael Bloomberg”, “Doctors Against Gun Violence”, “March for our Lives”, etc.

How to create a Boolean query (2)

- 5) Consolidate the terms.
 - Identify duplicates, synonyms, as well as any concepts that you want to combine even if they are not synonyms.
 - Re-label the term as needed to reflect the concept/category. Also consider and note any other relationships between terms. Prioritize the terms. Rank from 1-N, most relevant to least relevant.
 - Hint: Rank each term by higher, medium, lower relevance, then sort and rank from 1-7.
- 6) Write a query for each term.
 - Note that regular plurals (-s, -es, -ies) are usually (but not always) included automatically, but you always need to specify irregular plurals, e.g., “mice”.
- 7) Qualify the scope for each term.
 - Does the term require any qualification of the scope, e.g., by population, setting, geography, etc.?
 - Validate that the term is disjunctive, distinct, and requires no further qualification.
- 8) Combine the terms into a single nested query.

Example: Health Coverage – Brainstorm a list of 10 relevant words and phrases.

- ❖ Children's Health Insurance Program
- ❖ Eligibility and enrollment
- ❖ Health insurance
- ❖ Health plans
- ❖ Healthcare reform
- ❖ Insurance cost
- ❖ Medicaid
- ❖ Medicare
- ❖ Obamacare
- ❖ Uninsured

Example: Health Coverage –

Use that list to identify 10 relevant items (articles, videos, websites, etc.)

- ❖ State-by-State Coverage and Government Spending Implications of the Better Care Reconciliation Act and American Health Care Act
- ❖ The Impact of the AHCA on Federal and State Medicaid Spending and Medicaid Coverage
- ❖ How Repealing and Replacing the ACA Could Reduce Access to Mental Health and Substance Use Disorder Treatment and Parity Protections
- ❖ A Big Picture Look at the Individual Market, 2013-2017
- ❖ The Evidence on Recent Health Care Spending Growth and the Impact of the Affordable Care Act
- ❖ High-Risk Pools Under the AHCA
- ❖ Premium Tax Credits Tied to Age Versus Income and Available Premiums
- ❖ The Cost of Not Expanding Medicaid: An Updated Analysis
- ❖ Veterans Saw Broad Coverage Gains Between 2013 and 2015
- ❖ Ensuring Compliance with Network Adequacy Standards
- ❖ Recent Evidence on the ACA and Employment

Example: Health Coverage –

Review 10 relevant items and write down the words and phrases that provide a context for the theme/topic/concept.

- ❖ ACA
- ❖ Affordable Care Act
- ❖ AHCA
- ❖ American Health Care Act
- ❖ Better Care Reconciliation Act
- ❖ Expanding Medicaid
- ❖ Health insurance coverage
- ❖ Health insurance high-risk pools
- ❖ Health spending
- ❖ Individual health insurance market
- ❖ Insurance premium tax credits
- ❖ Insurance premiums
- ❖ Medicaid coverage
- ❖ Medicaid spending
- ❖ Mental health treatment
- ❖ Network adequacy standards
- ❖ Parity protections
- ❖ Substance use disorder treatment
- ❖ Uninsured
- ❖ Uninsured veterans
- ❖ Urban Institute
- ❖ Veterans' health insurance coverage

Example: Health Coverage –

Note any named entities (people, organizations, events, laws, etc.) that are closely associated with the theme/topic/concept.

- ❖ ACA
- ❖ Affordable Care Act
- ❖ AHCA
- ❖ American Health Care Act
- ❖ Better Care Reconciliation Act
- ❖ Children's Health Insurance Program
- ❖ Eligibility and enrollment
- ❖ Expanding Medicaid
- ❖ Health insurance
- ❖ Health insurance coverage
- ❖ Health insurance high-risk pools
- ❖ Health plans
- ❖ Health spending
- ❖ Healthcare reform
- ❖ Health insurance market
- ❖ Individual insurance premium tax credits
- ❖ Insurance premiums
- ❖ Medicaid
- ❖ Medicaid coverage
- ❖ Medicaid spending
- ❖ Medicare
- ❖ Mental health treatment
- ❖ Network adequacy standards
- ❖ Obamacare
- ❖ Parity protections
- ❖ Substance use disorder treatment
- ❖ Uninsured
- ❖ Uninsured veterans
- ❖ Urban Institute
- ❖ Veterans health insurance coverage

Example: Health Coverage – Consolidate the terms

Health insurance

- Eligibility and enrollment
- employer-sponsored insurance
- Federal facilitated marketplace
- Health insurance coverage
- Health insurance high-risk pools
- Health insurance exchange
- **HIE**
- **High deductible**
- Health insurance market
- Health plans
- **HMO**
- **Individual mandate**
- **In-network**
- Mental health treatment
- **Out-of-network**
- **PPO**
- **Pre-existing conditions**
- **Small business marketplace**
- **State marketplace**
- Substance use disorder treatment

Healthcare reform

- Health spending
- Individual insurance premium tax credits
- Insurance premiums
- Network adequacy standards
- Parity protections
- Uninsured

Affordable Care Act

- ACA
- American Health Care Act
- AHCA
- **BCRA**
- Better Care Reconciliation Act
- Obamacare

Children's Health Insurance Program

- CHIP

Medicare / Medicaid

- **Assistor**
- Expanding Medicaid
- Medicaid coverage
- Medicaid spending
- **Navigator**

Veterans health insurance coverage

- Uninsured veterans

Example: Health Coverage – Write a query for each term

Affordable Care Act

- ❖ ("Better Care Reconciliation Act" OR BCRA)
- ❖ ("American Health Care Act" OR AHCA)
- ❖ ("Affordable Care Act" OR "ACA" OR Obamacare)

CHIP

- ❖ ("CHIP" OR "Children's Health Insurance Program")

Medicare / Medicaid

- ❖ ((Medicare OR Medicaid) NEAR (spend* OR cover* OR expan*))
- ❖ ((navigator* OR assistor* OR assister*) NEAR (("health insurance" OR Medicare OR Medicaid) NEAR enroll*))

Veterans health insurance coverage

- ❖ uninsured NEAR veteran*

Example: Health Coverage – Write a query for each term

Health insurance

- ❖ ("health insurance coverage" OR "health coverage")
- ❖ ("health insurance exchange" OR "HIE")
- ❖ ("health insurance" NEAR market*)
- ❖ ("federal* facilitated market*" NEAR "health insurance")
- ❖ ((state NEAR/5 market*) NEAR "health insurance")
- ❖ ("small business market*" NEAR "health insurance")
- ❖ (("high-risk" OR "high risk") NEAR "health insurance")
- ❖ (("pre-existing condition*" OR "preexisting condition*") NEAR "health insurance")
- ❖ (("high-deductible" OR "high deductible") NEAR ("health insurance"))
- ❖ (("mental health" OR "substance abuse") NEAR "health insurance")
- ❖ ("provider network*" NEAR "health insurance")
- ❖ (("in-network" OR "out-of-network") NEAR "health insurance")
- ❖ ((PPO* OR HMO*) NEAR (market* OR plan* OR provider*))
- ❖ ("health insurance" NEAR (enroll* OR "re-enroll*" OR renew* OR "open-enrollment" OR "open enrollment"))
- ❖ ("individual mandate" NEAR "health insurance")
- ❖ "employer-sponsored insurance"
- ❖ ((employer OR employee) NEAR "health insurance")

Example: Health Coverage – Write a query for each term

Healthcare reform

- ❖ ("healthcare reform" OR "health care reform")
- ❖ (uninsured NEAR (veteran* OR child* OR adult* OR people OR kid* OR citizen*))
- ❖ "health insurance rate*"
- ❖ ((cost* OR rate* OR payment*) NEAR "health insurance")
- ❖ ("health insurance" NEAR "tax credit*")
- ❖ ((healthcare OR "health care") NEAR spending)

Example: Health Coverage – Combine the terms into a single nested query

- ❖ ("health insurance coverage" OR "health coverage") OR ("healthcare reform" OR "health care reform") OR ("Better Care Reconciliation Act" OR BCRA) OR ("American Health Care Act" OR AHCA) OR ("Affordable Care Act" OR "ACA" OR Obamacare) OR ((Medicare OR Medicaid) NEAR (spend* OR cover* OR expan*)) OR ("health insurance exchange" OR "HIE") OR ("health insurance" NEAR marketplace*) OR ("federal* facilitated marketplace*" NEAR "health insurance") OR ("federal* run marketplace*" NEAR "health insurance") OR ((state NEAR marketplace*) NEAR "health insurance") OR ("small business marketplace*" NEAR "health insurance") OR ("small-business marketplace*" NEAR "health insurance") OR (("small business" NEAR exchange*) NEAR "health insurance") OR (("high-risk" OR "high risk") NEAR "health insurance") OR (uninsured NEAR (veteran* OR child* OR adult* OR people OR kid* OR citizen*)) OR ("pre-existing condition*" OR "preexisting condition*") NEAR "health insurance") OR "health insurance rate*" OR ((cost* OR rate* OR payment*) NEAR "health insurance") OR ("health insurance" NEAR "tax credit*") OR ((healthcare OR "health care") NEAR spending) OR ((healthcare OR "health care") NEAR utilization) OR (("high-deductible" OR "high deductible") NEAR ("health insurance") OR (("mental health" OR "substance abuse") NEAR "health insurance") OR ("provider network*" NEAR "health insurance") OR (("in-network" OR "out-of-network") NEAR "health insurance") OR ((PPO* OR HMO*) NEAR (marketplace* OR plan* OR provider*)) OR ("health insurance" NEAR (enroll* OR "re-enroll*" OR renew* OR "open-enrollment" OR "open enrollment")) OR ((navigator* OR assistor* OR assister*) NEAR (("health insurance" OR Medicare OR Medicaid) NEAR enroll*)) OR ("CHIP" OR "Children's Health Insurance Program") OR ("individual mandate" NEAR "health insurance") OR "employer-sponsored insurance" OR ((employer OR employee) NEAR "health insurance"))

Summary

- ❖ Boolean queries are a useful way to scope custom categories.
- ❖ It is the most transparent method for specifying the rules for content categorizers to automate the tagging process to predefined categories or taxonomies.

Questions

Joseph Busch

jbusch@taxonomystrategies.com or joseph@semanticstaffing.com

@joebusch

mobile +1 415-377-7912

Boolean Query Business Rules: Summary

- ❖ There are lots of tools available that provide the building blocks for automated tagging applications including NLP, entity extraction, summarization, and sentiment analysis. Many tools and search engines also include a content categorizer. What they usually do is categorize ITPC news or Wikipedia categories. But what if you want to categorize to some other scheme or a set of custom subjects relevant to you or your organization's areas of interest? Boolean queries are a useful way to scope custom categories. It also happens to be the most transparent method for specifying the rules for content categorizers to automate the tagging process to predefined categories or taxonomies. This session provide a quick review of the Boolean query syntax, and then presents a step-by-step process for building a Boolean query categorizer.