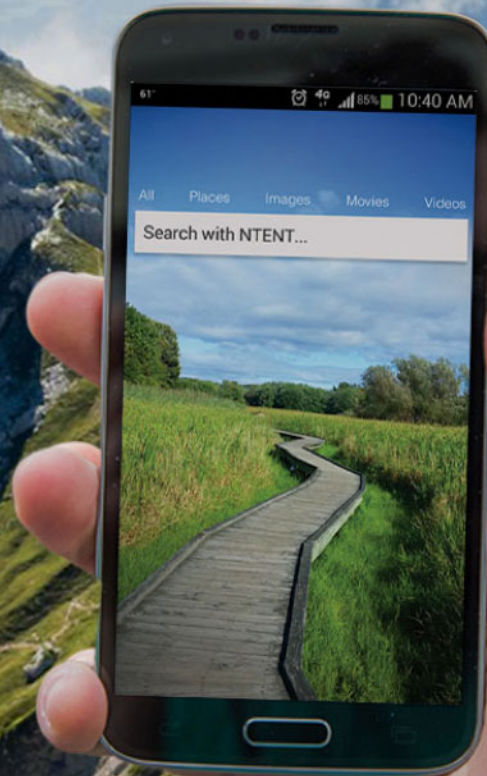


Semantic Mobile Search

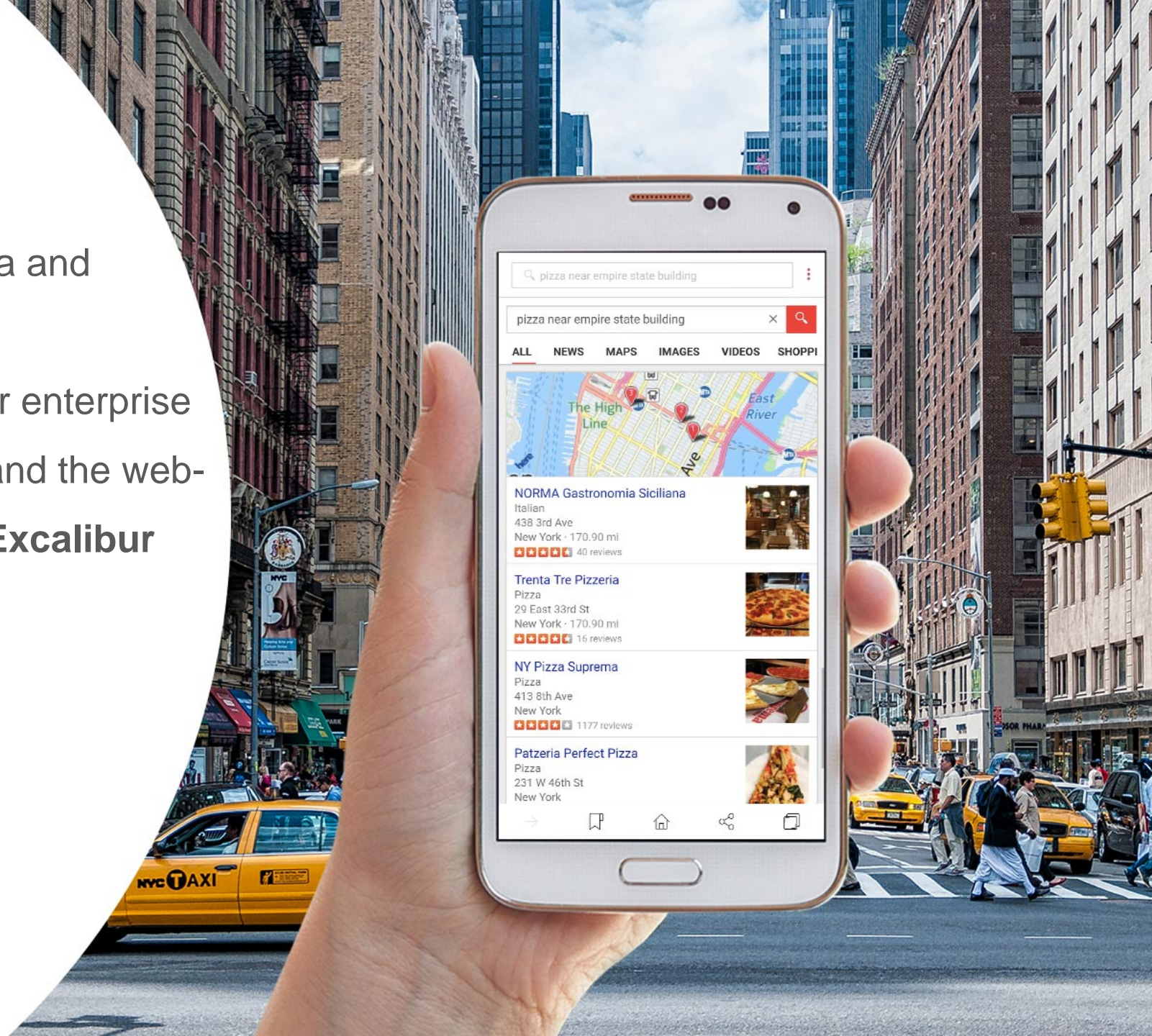
Andreas Kaltenbrunner

Senior Director of Data Analytics



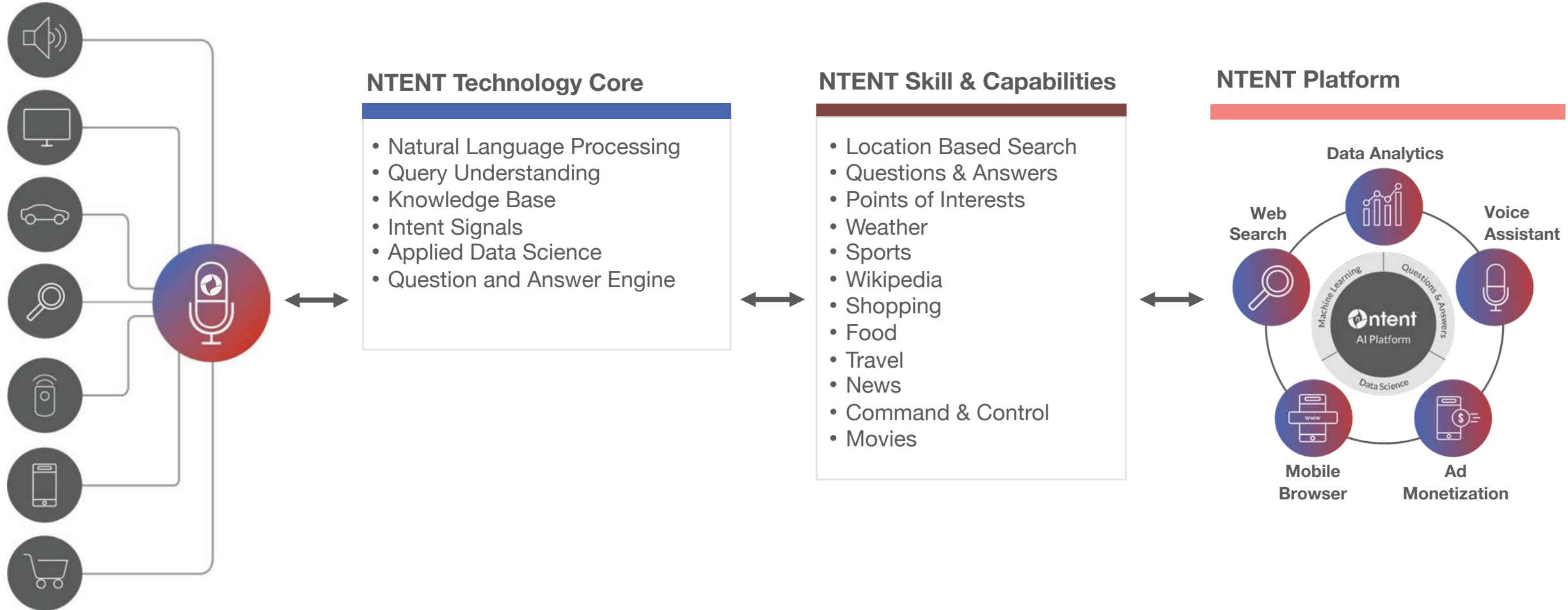
NTENT

- Semantic search technology
- Born from the merge of Convera and Firstlight ERA in 2010
- Convera is previously known for enterprise search engine **RetrievalWare** and the web-scale semantic search engine **Excalibur**
- Offices in:
 - New York, NY (HQ)
 - Carlsbad, CA (Engineering)
 - Barcelona, CAT (R&D)
 - Vienna, VA (USA)
 - London (UK)



At-A-Glance: NTENT's White-Label Platform

NTENT-Powered Applications





AGENDA

- Semantic Resources
- Document Processing
- Experts
- Query Understanding
- Usage Data Analysis



SEMANTIC RESOURCES

KNOWLEDGE RESOURCES

Composed of Four Parts

- Ontology (LI)
- Lexicon (LD)
- Onomasticon (mostly LI)
- Linguistic Rules (SemReps, Case Frames)

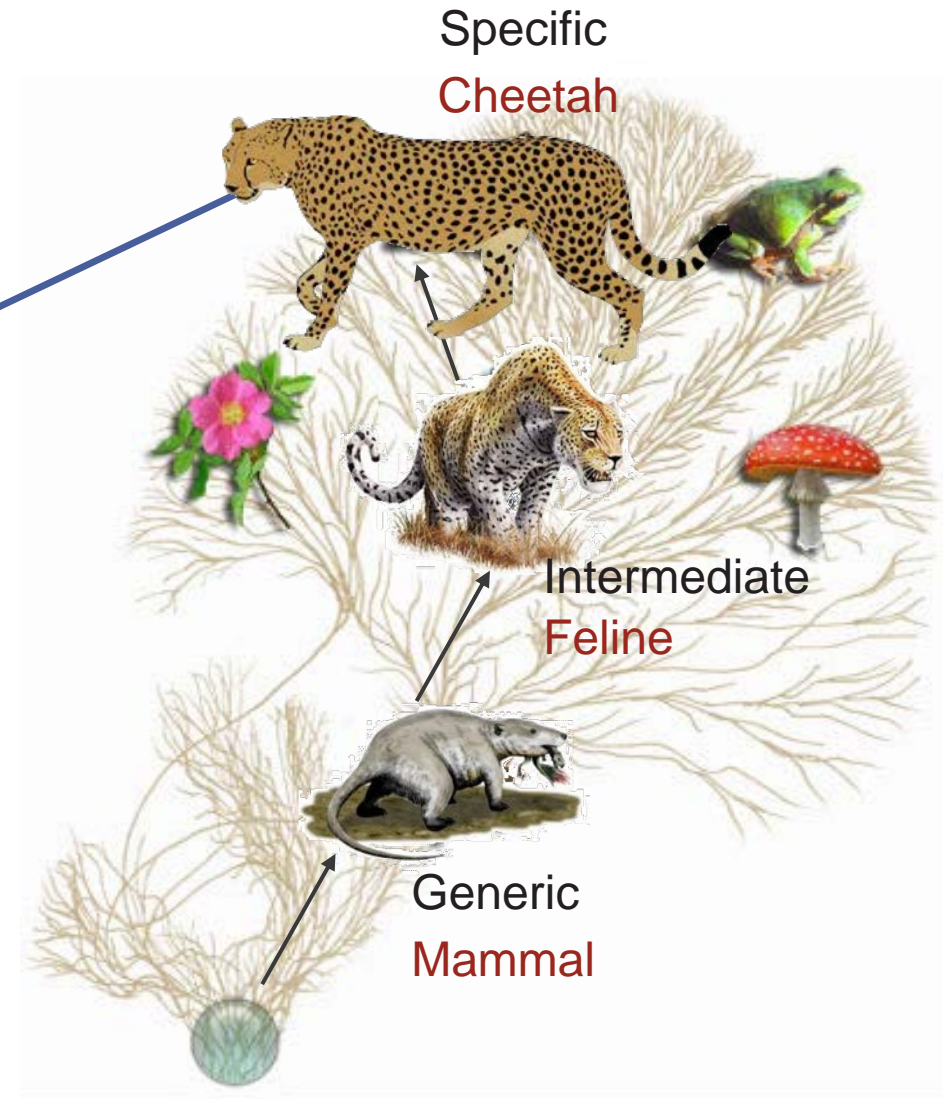


LEXICON

- Language **specific**
(one lexicon per language)
- One or more expression associated with each concept from the ontology
- Includes expression- specific attributes
- Current languages:
 - English
 - Russian
 - Turkish

en: cheetah
fr: guépard
es: güepardo

it: ghepardo
ru: гепард
cn: 獵豹



ONOMASTICON

- Generally Language **independent**
- One or more expression associated with each entity
- Includes expression-specific attributes
- Sometimes not language independent e.g. city of Geneva in Switzerland:
 - Geneva (en)
 - Genève (fr)
 - Ginebra (es)
 - Ginevra (it)
 - Genf (de)

en: Lady Gaga
fr: Lady Gaga
es: Lady Gaga



Generic
People



Intermediate
Singer

Specific
Lady Gaga



LINGUISTIC RULES

VARIOUS KINDS: CASE FRAMES, SEMREPS, ETC.

```
-- Given a context of lunch or dinner
-- and a cuisine is selected
-- returns list of restaurants
```


```
If
{
  Sequence
  {
    Or
    {
      Concept{"s:men.00BBL", desc = "dinner",
        match = "self" },
      Concept{"s:men.01A9B", desc = "lunch",
        match = "self" },
    },
  },
  Or
  {
```

Templates

```
{
  -- Covers questions such as "list of N ..."
  If
  {
    Sequence {
      Optional{ Text{"give me"} },
      Optional{ Or{ Text{"a"}, Text{"the"} }, },
      Text{"list"},
      Optional{ Text{"of"} },
      ZeroOrMore { Numbers_Entity, variables = "answer_count" },
    },
    Interpretation{
      name = "List of things", id="LISTOF",
      -- specify how many answers to fetch.
      FieldQuery{ name='Common.maxAnswerCount', Variable{'answer_count'
    },
  },
},
```

Templates

```
{
  VerbFrame
  {
    lex="buy",
    desc="to purchase goods or services",
    sem="s:buyEvent", sem_desc="buy event",
    Syn{
      ex="Mary buys [pizza]", id="1",
      NP{role="agent", role_range="s:kec.0051B,s:cvc.CCEDN", direction="-"},
      VP{"buy"}, Optional{NP{role="theme", role_range="s:gen.002T5,s:gen.002W5,s:nasa.00AOP,s:fctw.inanprp"},},
    },
    Syn{
      ex="__ buy pizza", id="2",
      VP{"buy"}, NP{role="theme", role_range="s:cpr.00411,s:kec.C9TL0,s:gen.002W4"},
    },
    Syn{
      ex="Chris bought [donuts] for brunch.", id="3",
      NP{role="agent", role_range="s:kec.0051B,s:cvc.CCEDN", direction="-"},
      VP{"buy"}, Optional{NP{role="theme", role_range="s:gen.002T5,s:gen.002W5,s:nasa.00AOP,s:fctw.inanprp"},}, PP{"for", role="goal", role_range="s:gen.00007,s:fctw.inanprp"},
    },
    Syn{
      ex="__ bought [donuts] for brunch.", id="4",
      VP{"buy"}, Optional{NP{role="theme", role_range="s:gen.002T5,s:gen.002W5,s:nasa.00AOP,s:fctw.inanprp"},}, PP{"for", role="goal", role_range="s:gen.00007,s:fctw.inanprp"},
    },
    Syn{
      ex="__ pizza buy", id="5",
      NP{role="theme", role_range="s:gen.002W5,s:gen.002T5"}, VP{"buy"},
    },
    Syn{
      ex="__bought toys for children", id="6",
      VP{"buy"}, Optional{NP{role="theme",role_range="s:gen.002T5,s:gen.002W5,s:nasa.00AOP,s:fctw.inanprp"},}, PP{"for", role="benef",role_range="s:kec.0051B,s:cvc.CCEDN,s:men.006D1,s:men.D000829"},
    },
  },
},
```

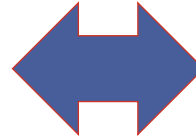
A large, irregular red ink splatter or blotch is centered on a white background. The splatter has a textured, watercolor-like appearance with various shades of red and some darker, almost black, areas. The edges are feathered and there are smaller droplets or splatters radiating outwards from the main shape.

DOCUMENT PROCESSING

SEMANTIC ENGINE



Semantic Engine



Ontology / Lexicon / Onomasticon


Concepts (common nouns)
Entities (proper names)


SEMANTIC ENGINE FUNCTIONS

- Language detection
- Boilerplate detection and removal
- Tokenization and lemmatization
- Part of speech tagging
- Morphological analysis
- Entity extraction and general dictionary matching
- **Concepts identification and disambiguation**
- Document scoring and classification









EXAMPLE: WEB PAGE

 **USA TODAY**
A GANNETT COMPANY


Search 


SUBSCRIBE NOW
3 MONTHS FOR
\$19.95

NEWS SPORTS LIFE MONEY TECH TRAVEL OPINION 68° CROSSWORDS YOUR TAKE INVESTIGATIONS VIDEO STOCKS APPS MORE 


 10
 76




Rockets can't afford to have Dwight Howard slowed by injury

 **Sam Amick**, USA TODAY Sports 6:38 p.m. EDT May 20, 2015

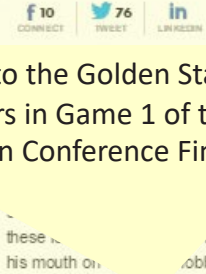


USATODAY Sports' Sam Amick breaks down Golden State's win to kick off the Western Conference finals.



... loss to the Golden State Warriors in Game 1 of the Western Conference Finals.

(Photo: Kyle Terada, USA TODAY Sports)



... loss to the Golden State Warriors in Game 1 of the Western Conference Finals.

... in Houston two ... playing in games like these ... the worst kind of taste in his mouth on ... obbled by a left knee bruise that kept him conta ... ed through most of the Rockets' 110-106 loss to the Golden State Warriors in Game 1 of the Western Conference Finals. Behold your early X-factor in this series: The Rockets aren't likely

One Question Site Survey

IT TAKES ONLY SECONDS TO ANSWER BELOW

How likely are you to watch NBA games during Playoffs?

SELECT ONE ANSWER

☐ Definitely will

☐ Probably will

☐ May or may not


☐ Probably will not

☐ Definitely will not

VOTE TO SEE RESULTS

POWERED BY VIZU [SAFE & ANONYMOUS](#)

RELATED VIDEO



NBA mock draft 8.0: Who should be the top 5 picks?

ENTITY EXTRACTION

Attested Entities

- Named Entities: people, places, organizations, brands, works of art, etc.
 - Represented in the ontology with relations to other concepts

Unattested (not in the Onomasticon)

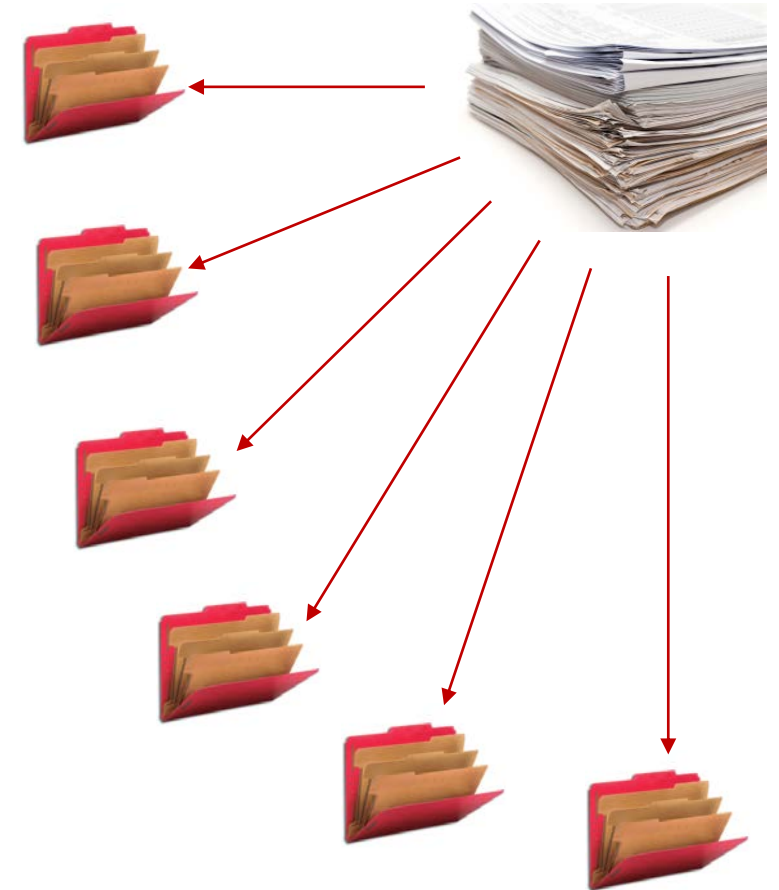
- Named Entities: people, places, organizations, etc.
 - Inferred people, places, organizations, etc. based on clues (e.g., “Mayor”, “Hospital”, “Inc. “, “Mr.”)
 - Machine learning based NER
- Other Entity Types: phones, monies, date/time, etc.
 - Pattern based, localized

IDENTIFY DOCUMENT CHARACTERISTICS

Uses rule-based and machine-learning classification

Example: If document mentions a sports team or an athlete, then classify as sports-category

- Identify broad document topic (e.g., health, politics)
- Identify aspect around a topic (e.g., entertainment, places to stay, restaurant, weather)
- Identify document genre (biography, FAQ, contract, consumer guides, recipes, analyst opinion)
- Identify spam, adult, offensive content, smoking, weapon use, gambling, unfortunate events
- Identify redaction quality (scientific article, conversation)
- Identify special document characteristics (many links, typos, shopping cart, login)



EXAMPLE: IAB category classifier

Classification of queries into predefined categories

IAB (Interactive Advertising Bureau) topic taxonomy

<https://www.iab.com/guidelines/taxonomy/>

- Takes advantage of rich, pre-trained word embedding models that leverage large corpora.
- No need to annotate large amount of queries
- Instead, generate suitable keywords for each of the categories.
- Predict the category in the following way:
 1. Generating a 'canonical vector', encodes the overall semantic properties across the keywords of a category.
 2. Assign an uncategorized query to closest 'canonical vector',

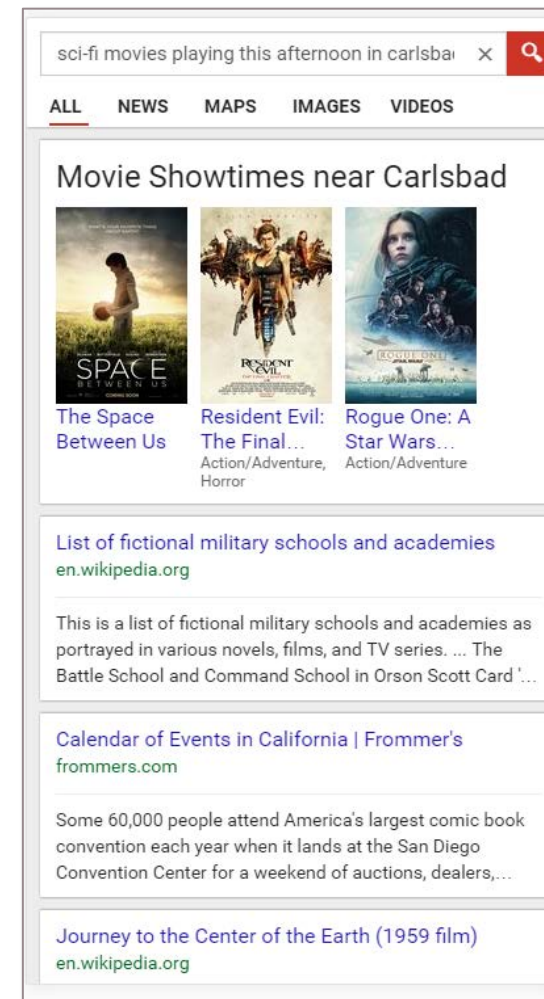
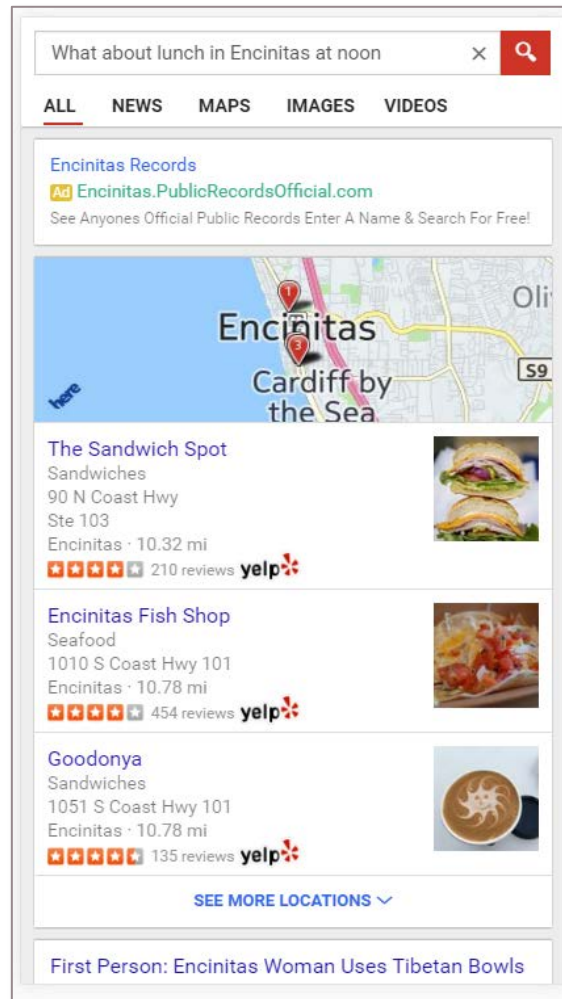
Category_cohort	2018_September	2019_January
Technology and Computing	13,4%	14,9%
Pop Culture	9,7%	11,9%
Travel	4,8%	6,3%
News and Politics	7,5%	6,2%
Television	4,7%	5,3%
Sports	2,6%	3,9%
Shopping	3,2%	3,5%
Events and Attractions	2,5%	3,5%
Home and Garden	2,8%	3,5%
Food and Drink	3,2%	3,3%
Movies	4,7%	3,2%
Family and Relationships	2,3%	2,5%
Education	2,7%	2,4%
Automotive	1,9%	2,4%
Religion and Spirituality	2,3%	2,3%
Pets	1,5%	2,2%
Healthy Living	1,9%	2,0%
Video Gaming	1,3%	2,0%
ADULT	10,5%	1,9%
Fine Art	1,4%	1,8%
Business and Finance / Business	1,8%	1,8%
Business and Finance / Economy	2,5%	1,7%
Science	1,3%	1,6%
Music and Audio	1,5%	1,6%
Business and Finance / Industries	1,5%	1,5%
Style and Fashion	1,5%	1,4%
Careers	1,3%	1,3%
Medical Health	1,0%	1,2%
Real Estate	0,9%	0,9%
Hobbies and Interests	0,8%	0,9%
Books and Literature	0,6%	0,7%
Personal Finance	0,7%	0,7%
TOTAL	100,0%	100,0%

Impact of an adult filter on query category distribution



EXPERTS

MOBILE SEARCH RESULTS - BEYOND TEN BLUE LINKS



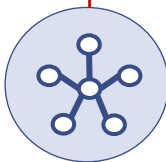
APPLICATION: EXPERT ANSWERS

Query

PF Chang near Encinitas

Formal Interpretation

```
1. Interpretation for 'MEAL_PLAN' (rank:1.14167)
  1. Situation (id:MEAL_PLAN)
    1. Action (id:RESTAURANT_ACTION_EST, name:Let's go out for a meal, model:null,
      rank:1.25)
      MEAL_PLAN access P.F. Chang's Encinitas, CA
      • FieldValue: Query.Class = local (null)
      • FieldValue: Query.Topic = Foodstuff Vertical (vt:vrt.CD5W9)
      • FieldValue: Directory.Business.generalCategory = restaurant (s:pct.001CR)
      • FieldValue: Directory.Business.name = P.F. Chang's (s:kec.00KZG)
      • FieldValue: Common.Location.city = Encinitas, CA (s:usg.1652705)
      • FieldValue: Common.Location.region = California (s:usg.0008D)
      • FieldValue: Common.Location.regionCode = CA (null)
      • FieldValue: Common.Location.country = United States of
        America (s:geocn.000JQ)
      • FieldValue: Common.Location.countryCode = US (null)
      • FieldValue: Common.Location.lat = 33.037 (null)
      • FieldValue: Common.Location.long = -117.292 (null)
```



...

EXPERTS – BEST SOURCES OF KNOWLEDGE

A single authority on a particular type of result or subject type. Examples:

Yelp = Local Business Listings

Stats.com = Major league sports data

NTENT treats its own ingested results as experts:

NTENT news and web-based search index

Three flavors of expert:

External API call – an external API is called that resides outside of NTENT data centers

Ingested data from third parties – offline or batch data is regularly ingested

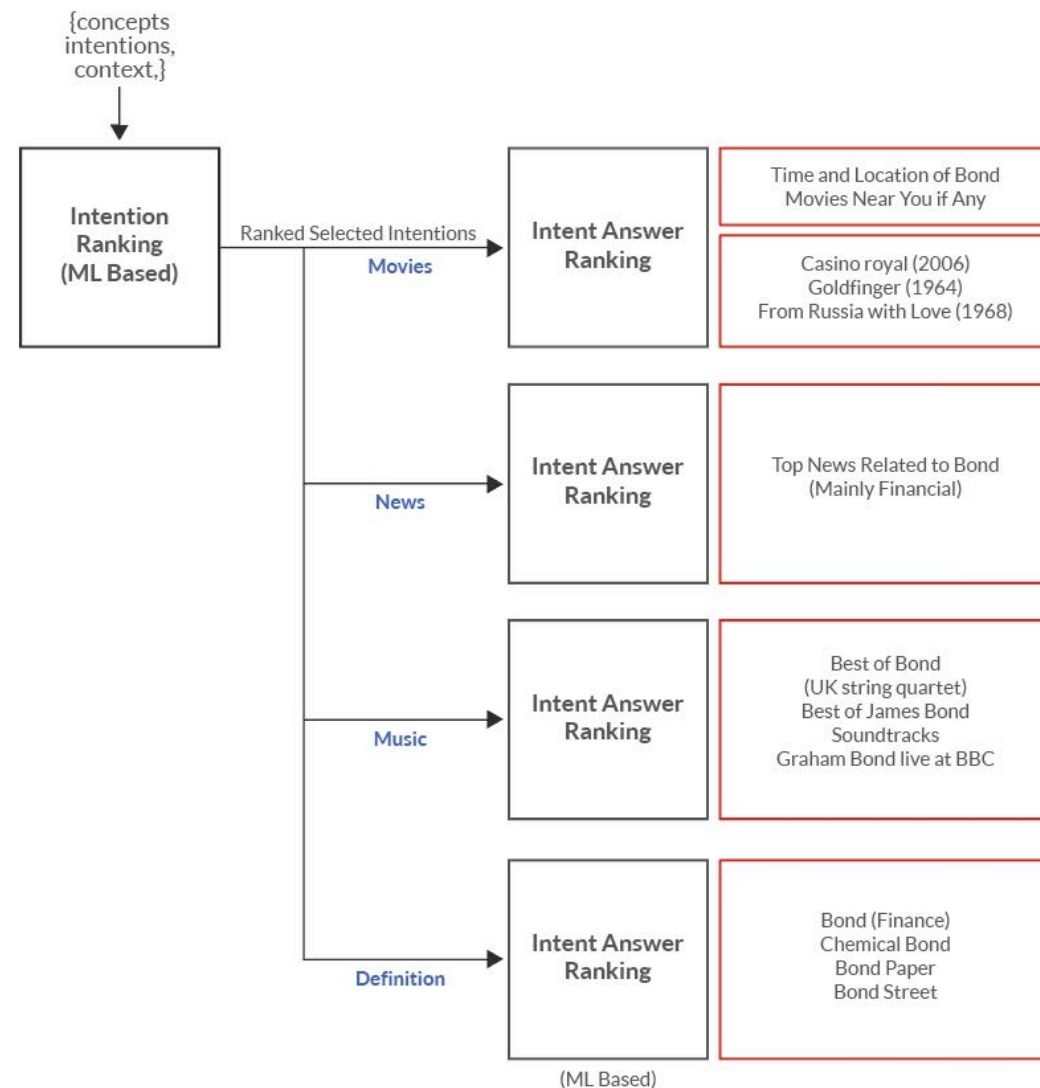
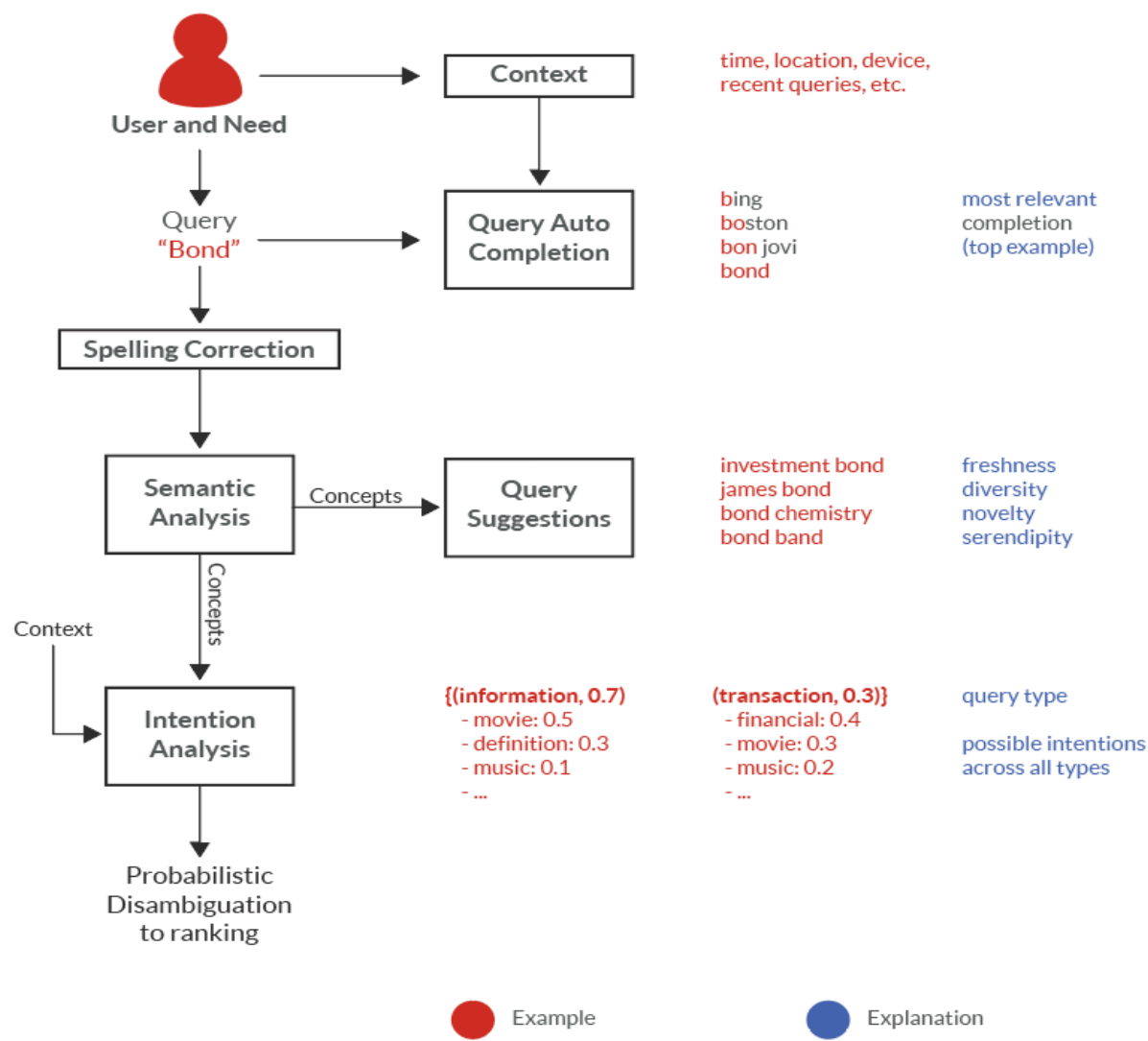
NTENT's own expert data



A large, irregular red ink splatter or blotch is centered on a white background. The splatter has a dark red core that fades into a lighter red and then into the white background, with many small droplets and specks trailing outwards. The text 'QUERY UNDERSTANDING' is written in white, uppercase, sans-serif font across the center of the dark red area.

QUERY UNDERSTANDING

QUERY UNDERSTANDING WORKFLOW

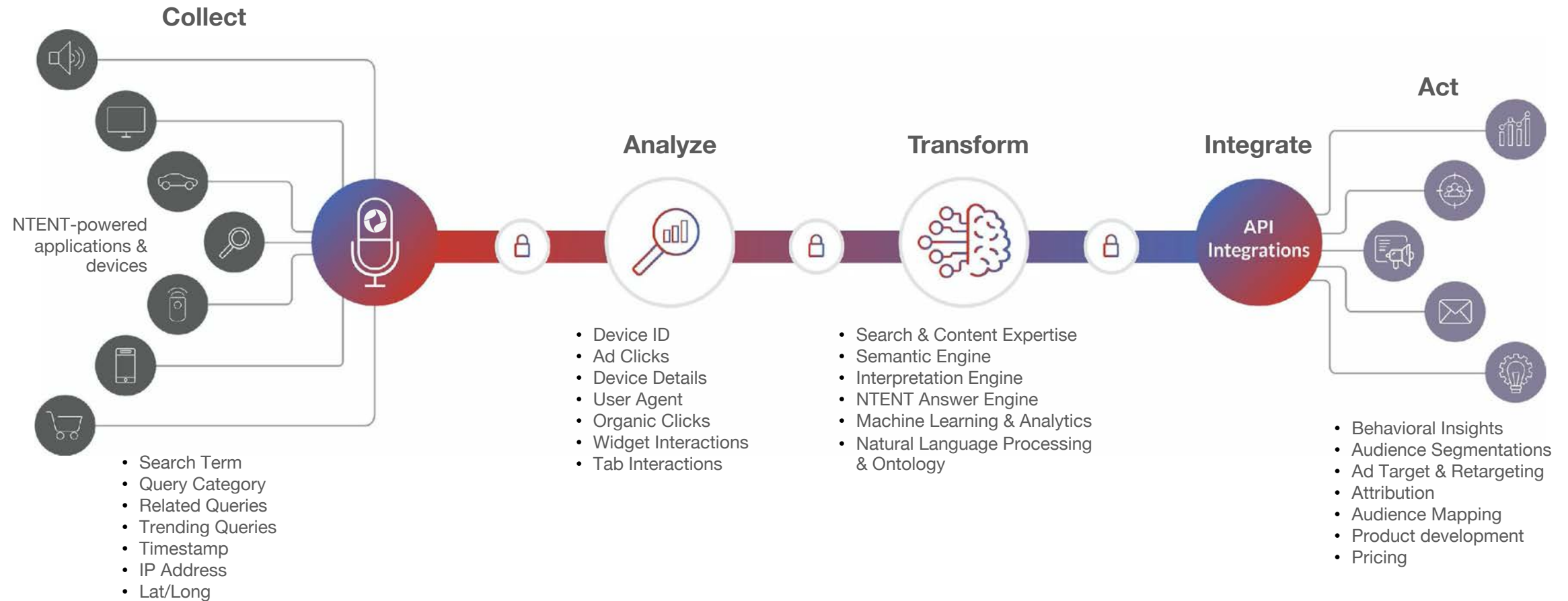


A large, irregular red ink splatter or blotch is centered on a white background. The splatter has a textured, watercolor-like appearance with various shades of red and some darker, almost black, areas. The edges are fuzzy and spread out, with smaller droplets and splatters visible around the main shape.

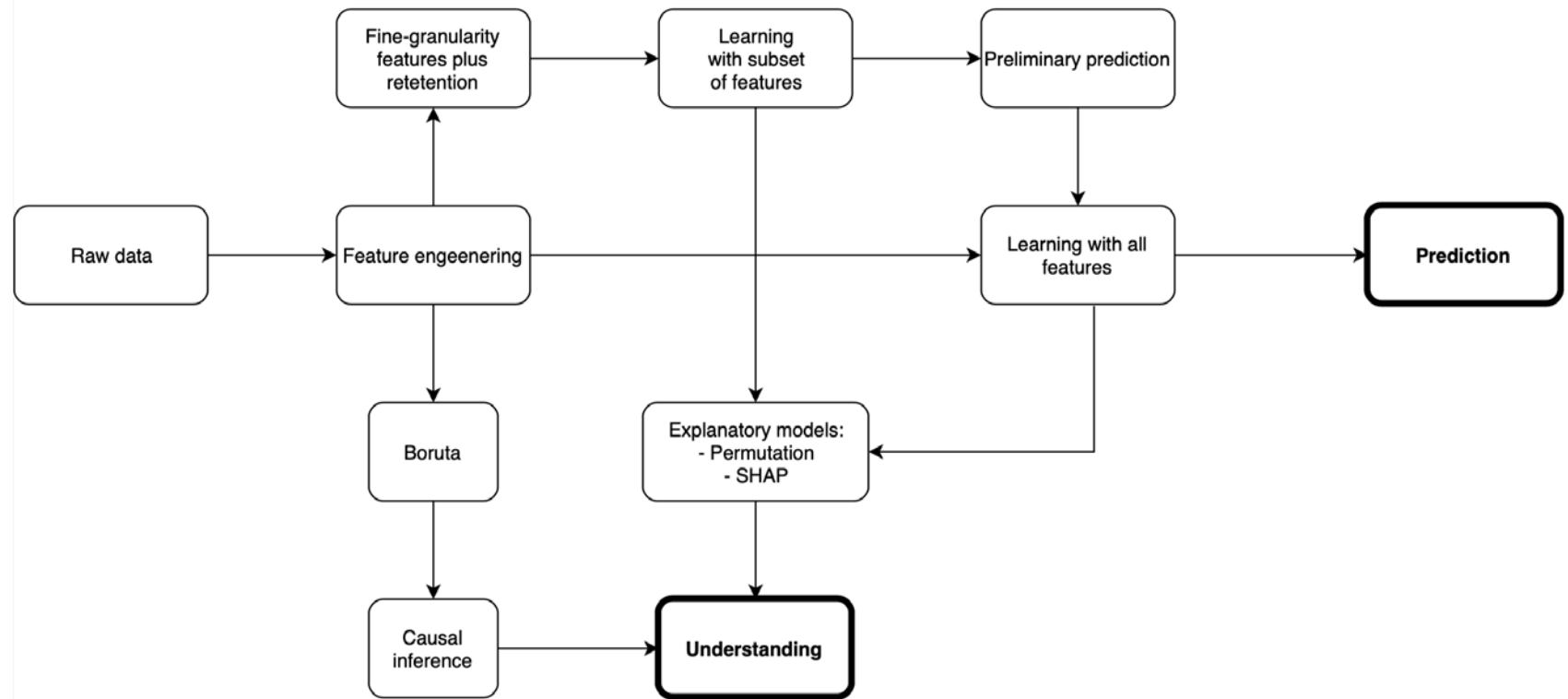
USAGE DATA ANALYSIS

NTENT Data Analysis, Transformation and Activation

Our deterministic data is very valuable, can be monetized and used to enhance digital marketing and product development.



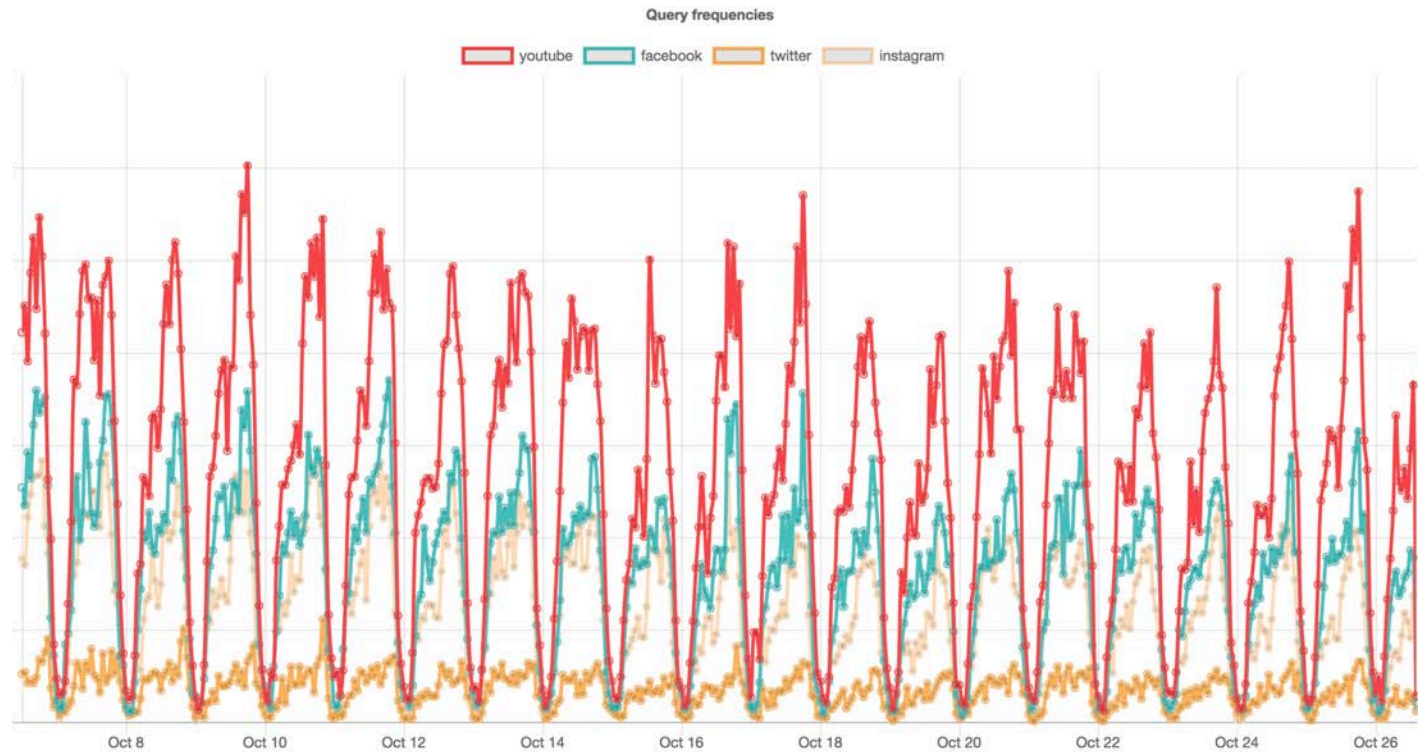
Example: Churn and Activity Prediction



Objectives:

- Create classifiers to predict user activity and possible user churn.
- Detect positive or negative impact of certain features on user retention

Example: Real Time Query Analysis



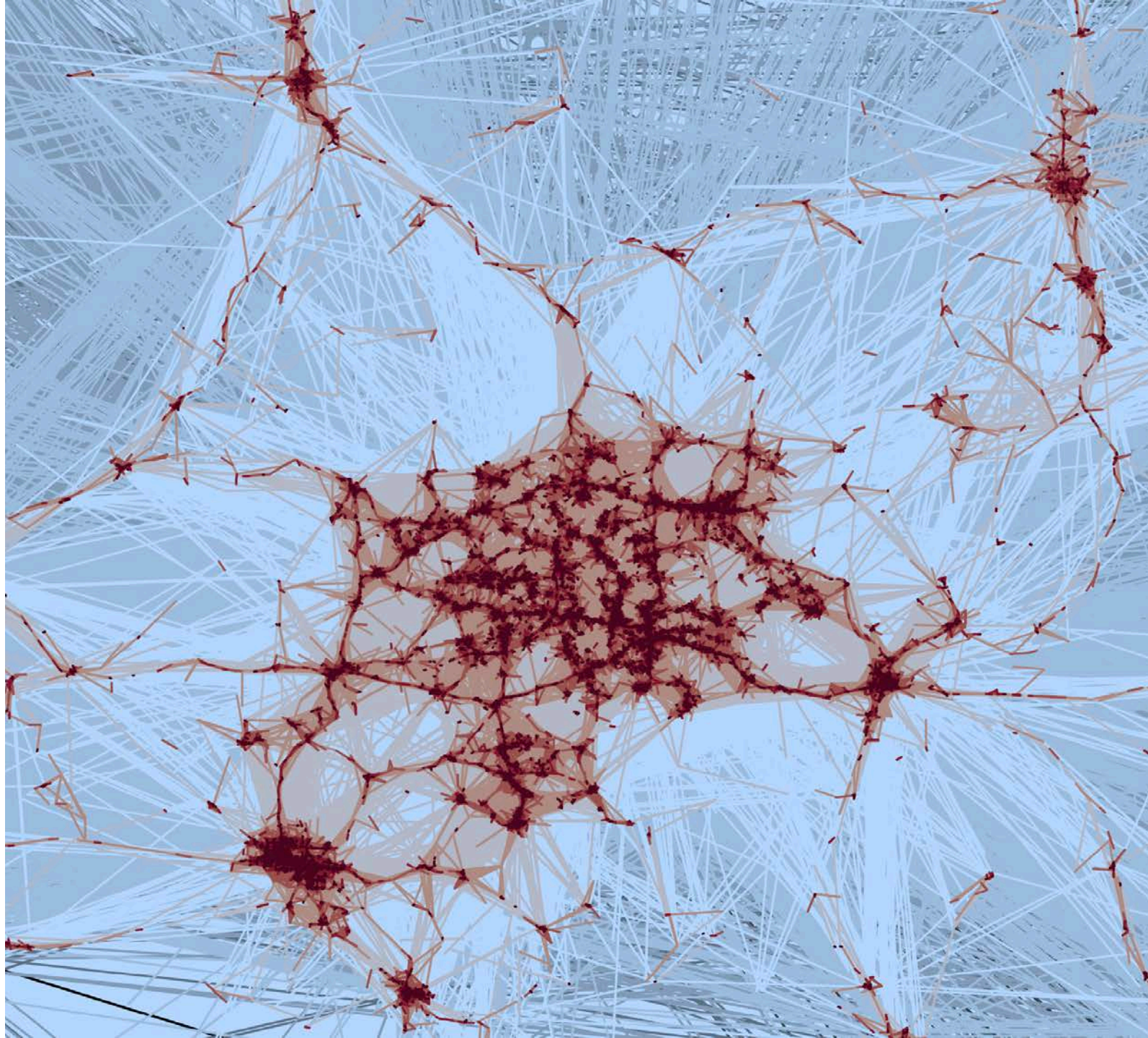
Objectives:

- Real time monitoring and visualization systems for search traffic analysis.
- Detect trends and incidents

Example: Mobile Search User Movements

Objectives:

- Improve search results through location and trajectory analysis.
- Target advertising based on movement data.



Questions?



ntent[™]

Smart Search. Fast Answers.

www.ntent.com | [@withntent](https://twitter.com/withntent) | 877.861.2230