

Beyond Text: Searching Rich Media

Sue Feldman

Search and Discovery Technologies

IDC

Agenda

- Elements of search
- Matching techniques and technologies
- Features to extract
- Uses
- Questions and issues

Elements of Search

- Varied content, multiple sources, few formal requirements for input structure
- Feature extraction
- Document representations based on vectors, graphs, etc.
- Similarity (fuzzy) matching
- Architecture for ad hoc retrieval
- Natural (human) interaction for query and results browsing

Matching techniques

Similar but not the same features determined by:

- Measuring degree of similarity between query and document representations
- NLP—syntactic and semantic matching
- Pattern matching
- Prediction of probability of a match
- Context of the query elements
- Categorization,
- Machine learning, with or without training sets
- Tagging
- Usage and users
- Voting algorithms
- Multiple sources of evidence
- Dictionaries, vocabularies, taxonomies, ontologies

Types of features

- Words, phrases
- Phones, phonemes
- Concepts and relationships among them
- Intonation patterns
- Facial features
- Colors
- Shapes
- Lines
- Scene changes
- Fingerprint whorls

Uses

- Find images, videos by topic or feature
- Spam detection
- Intellectual property protection
- Security monitoring (faces, actions)
- Law enforcement (transcribe speech, match fingerprints, faces)
- Dating sites (find me other guys who look like this...)
- Medical detection (suspicious shapes), telemedicine
- Call centers (identify angry customers)
- Emergency response teams
- Ad matching

Questions and issues

- Which features are important for each application?
- How will users ask for content (Sunset? Vacation? Beach? Quercus? Nathan Milstein? Violinists?)
- How to query non-text content?
- Scalability and speed
- Dynamic collections
- What does an emergency sound like?
- What does happiness look like?

Contact Information



Susan Feldman
sfeldman@idc.com

- Sue Feldman, VP, Search and Discovery Technologies, IDC
- Tom Wilde, CEO, Everyzing
- Naveen Agnihotri, CTO, Milabra
- Mike Phillips, CTO, and co-founder, Vlingo