

# Index

## A

Ablation-style evaluation, 72  
Abstract properties, 16  
Alexa, 87  
Alias, 20  
Analogy, 21, 61  
Arguments, 6, 23  
Assertion, 2  
AUC-PR, 73  
Automated ER solutions, 52  
Automatically generated discriminator phrases, 16  
Automatic content extraction (ACE), 22  
Automation-scalability tradeoff, 54

## B

Background corpus, 17  
Background knowledge, 35  
Bags of words, 18  
Baseline statistical learning, 13  
Baseline system, 13  
Bayesian optimality, 52  
BBN ACCENT, 37  
Big Data, 1  
Bi-grams, 18  
Bilateral pairs, 38  
Blocking, 39  
    key, 40  
    learning, 43  
    quality, 43  
    value, 40  
    method, 40  
    without supervision, 56

Block purging, 40  
Bootstrapping, 14

## C

Candidate entity, 16  
Candidate set, 44  
    reduction, 48  
Canopies, 42  
Capitalization feature, 16  
Case citations, 14  
CBOW, 21  
Character makeup, 17  
Character n-gram, 17  
Chunk, 18  
Clerical review, 52  
Clustering, 15  
COALA, 53  
Common corpus, 21  
Common nouns, 19  
Commonsense reasoning, 1  
Conditional random fields, 10  
Context, 20, 21, 61  
Contextual clues, 14  
Coreference resolution, 12, 20  
Cortana, 87  
Cosine similarity, 18  
Cross-document coreference resolution, 33  
Cue words, 19

## D

Data ecosystems, 36  
Data skew, 36, 40, 43

DBpedia, 18, 78, 80  
 Dedoop, 54  
 Deduplication, 34  
 Deep neural networks, 16  
 DeepWalk, 62  
 Derivational suffixes, 19  
 Diacritics, 19  
 Dictionary, 18  
 Disambiguation, 19  
   rules, 13  
 Discriminative features, 13  
 Disjunctive normal form, 43  
 Distributional similarity, 14  
 DKRL, 69  
 Document-centric features, 19  
 Document and Corpus Features, 19  
 Domain adaptation, 37  
 Domain engineering, 45  
 Domain-specific KGC, 10  
 Downstream NLP, 61  
 Dublin Core, 79  
 Duplicated coverage, 48

## E

Edit-distance, 19  
 Energy-based embedding model, 63  
 Entity classification, 73  
 Entity Resolution (ER), 6, 33  
   automation, 36  
   heterogeneity, 36  
   related work, 51  
   scalability, 36  
 Entity typing, 18  
 EUCLID, 53  
 Events, 6  
   domain, 4  
   extraction, 11, 24  
 Evolution of ER, 45  
 Exhaustive set, 49  
 Expectation maximization, 52  
 Extending the two-step workflow, 51  
 External lexicon, 18

## F

Feature-based blocking, 42  
 Features, 16  
   engineering, 16, 20, 61  
   function, 17  
   library, 45  
   vector, 45  
 Fellegi-Sunter method, 44, 52  
 Finer-grained typing, 18

First-order entities, 6  
 Fixed-size window, 15  
 F-measure, 50  
 Freebase, 3  
 Fuzzy matching, 19

## G

Gazetteer, 18  
 General AI, 1  
 Generalization, 13, 14, 66  
   capabilities, 16  
 Genetic ER algorithms, 53  
 Geopolitical events, 37  
 GKV, 82  
 Global data space, 76  
 Global loss function, 63  
 Glove, 21  
 Google Knowledge Graph, 1  
 Google Knowledge Vault, 82  
 Google news corpus, 21  
 Graph-aware embedding models, 71  
 Graph embeddings, 38  
 Graph priors, 82  
 Greedy algorithms, 44  
 Ground-truth, 44

## H

Hadoop, 41  
 Head entity, 2, 62  
 Heuristics, 45  
 Hierarchical graphical models, 52  
 HTML, 11  
 Human in the loop, 14  
 Hungarian algorithm, 47  
 Hyperparameter tuning, 66  
 Hyponyms/hypernyms, 15

## I

I.i.d, 44  
 In-KG applications, 72  
 Inflectional, 19  
 Infobox, 80  
 Information extraction, 9  
 Information retrieval, 15  
 Instance-based blocking, 42  
 Internationalized resource identifiers (IRI), 5

## J

Jaro-Winkler, 19  
 Joint IE, 25

Joint modeling, 24

Joint text-video extraction, 10

JSON-LD, 86

## K

Knowledge bases, 3, 4

Knowledge graph

adoption, 75

completion, 59

ecosystems, 75

identification, 59

Knowledge graph embeddings (KGEs), 60

applications, 72

Knowledge panels, 1

Knowledge repository, 82

Knowledge sub-model, 68

Knowledge Vault, 82

## L

Labeled training data, 10

Latent Dirichlet allocation (LDA), 52, 61

Lemmatized, 19

Less well-known entities, 19

Lexical resources, 15

Lexicon, 18

LIMES, 54

Linguistic patterns, 14

Linked Data, 76

principles, 77

technology stack, 78

Linking Open Data (LOD), 77

Link prediction, 72

Link specification function, 38

List, 18

List-lookup features, 18

Literal denotation, 20

Literals, 9

Locality sensitive hashing (LSH), 54

Long short-term memory (LSTM), 16

Low-dimensional, 20

## M

Machine learning effectiveness, 20

MapReduce, 41, 54

Markup, 9

Mean rank, 73

Mean reciprocal rank, 73

Measuring blocking performance, 48

Merge purge, 41

Message Understanding Conferences (MUCs),  
10

Meta-ability, 37

Metadata, 20

Meta-information, 19

Metaphone, 19

Metonymy, 20

Microformats, 86

Minimal supervision, 53

Monge-elkan, 47

Morphological feature, 17

MUC-6 training, 13

Multi-lingual, 17

Multi-pass sorted neighborhood, 42

Multi-relational graph, 62

Multi-word expression, 19

Mutual bootstrapping, 14

## N

Named entity detection, 18

Named entity recognition (NER), 11, 12

N-ary, 22

Natural languages, 9

NLP-centric IE, 11

Nominal feature, 16

Normalizing, 19

NTN, 68

NYSIIS, 46

## O

Ontological information, 67

Ontology matching, 56

OOV, 21

OpenCyc, 86

Open IE, 10, 13

Open Knowledge Network (OKN), 86

Out-of-KG, 69

Out-of-KG applications, 74

## P

Pairs completeness, 49

Pairs quality, 49

Patient linking, 34

Pattern features, 18

Pattern generalization, 14

PC-RR tradeoff, 49

PFM, 53

Phonetic algorithm, 19

Phonetic code, 19

PMI-IR, 15

Pointwise mutual information, 15

Post-processing steps, 51

Precision, 15, 50, 73

Probabilistic soft logic (PSL), 66  
 Product matching, 37  
 Products and e-commerce, 5  
 Projection matrix, 65  
 Property heterogeneity, 37  
 Property matching, 51  
 Publication, 4, 39

## Q

Quadratic complexity, 38  
 Quantitative feature, 16

## R

Rare named entities, 15  
 RAVEN, 55  
 RDF, 62  
 RDFS, 79  
 Real-world ER, 46  
 Recall, 50, 73  
 Recall-friendly, 51  
 Receiver operating characteristic (ROC), 50  
 Red-blue set covering, 44  
 Reduction ratio, 49  
 Reference, 19  
 Relation-specific hyperplanes, 64  
 Relational features, 20  
 Relational information, 53  
 Relation Extraction (RE), 11, 22  
 Relation path, 71  
 Relationship, 2  
 Representation learning, 16, 38, 61  
 RoadRunner, 28  
 Robust, 16  
 Rule base, 45  
 Rules and KGEs, 69

## S

Same semantic class, 14  
 Schema.org, 84  
 Schema heterogeneity, 51  
 Second-order entities, 6  
 Seed entity, 14  
 Segmented noun phrases, 16  
 Selecting the unlabeled data, 15  
 Semantic class, 21  
 Semantic dependence, 61  
 Semantic regularity, 60  
 Semantic relation, 22  
 Semantic tagging, 20  
 Semantic Web, 3, 33, 62, 76  
 Semi-supervised learning, 14

Sequence-labeling, 16  
 SILK, 54  
 Similarity, 44  
   of context, 15  
 Site generation, 28  
 Skip-gram, 21  
 SKOS, 79  
 Sliding window, 41  
 Slot fillers, 9  
 Sorted neighborhood, 41  
 Soundex, 19  
 SPARQL, 77  
 Special character, 17  
 Spelling errors, 18  
 SSE, 67  
 Stable marriage, 53  
 Stemmed, 19  
 String similarities, 19  
 Structural homogeneity, 38  
 Structured data, 1  
 Supervised machine learning, 10, 13  
 Swoosh, 54  
 Synonyms, 14  
 Synset, 15  
 Syntactic units, 16

## T

Tail entity, 2, 62  
 Temporal information and KGEs, 70  
 Terrorist attack events, 24  
 Textual description incorporation into KGE  
   model, 68  
 Tf-idf, 18, 20, 46, 52, 61  
 Thresholded edit-distance, 19  
 Time-aware embedding, 70  
 TKRL, 67  
 Token-based, 18  
 Traditional blocking, 40  
 Trans\*, 64  
 TransE, 63  
 TransE extensions, 64  
 Transfer learning, 53  
 TransH, 64  
 Translations, 63  
 TransR, 64  
 Tri-grams, 18  
 Triple, 2  
 Triple classification, 73  
 Triplify, 10  
 Twitter, 14  
 Two-layer similarity feature, 46  
 Two-step framework, 38  
 Type-specific projection matrix, 67

Type heterogeneity, [36](#)

Type hierarchy, [37](#)

Type matching, [51](#)

Type signatures, [15](#)

## U

Underlying ontology, [4](#)

Unlabeled corpora, [38](#)

Unlabeled corpus, [15](#)

Unsupervised learning, [14](#)

Unsupervised machine learning, [15](#)

URL, [77](#)

## V

Vector space, [61](#)

Viable linking candidates, [35](#)

Vocabulary transfer, [13](#)

## W

Weakly supervised, [14](#)

Web IE, [11](#)

Web of data, [75](#)

Web of things, [76](#)

Web queries, [15](#)

Wikidata, [3](#), [86](#)

Wikipedia, [3](#), [21](#)

Wikipedia anchors, [69](#)

Word-level features, [17](#)

Word2vec, [21](#), [61](#)

Word embeddings, [20](#)

WordNet, [15](#)

Word polysemy, [18](#)

## Y

YAGO, [18](#), [67](#)