

Index

- Alignment Technique
 - single document summarization, 26
- Artificial Intelligence, 17
- Automatic Speech Recognizer, 4, 24, 33, 34
 - acoustic feature, 34
 - discourse feature, 34
 - JANUS, 125, 127, 136
 - lexical feature, 34
 - listener feedback, 34
 - non-textual input, 4, 48
 - prosodic feature, 34
 - read text, 34
 - speaker activity, 34
 - spontaneous speech, 34
 - textual output, 24, 48
 - transcript, 34
 - ViaVoice, 127
 - Word Error Rate, 125, 127
- Automatic Summarization, 2, 11, 15
 - definition, 2
 - evaluation, 6
 - input aspect, 2
 - input aspects, 48
 - lexical chain, 69
 - output aspect, 5
 - output aspects, 48
 - purpose aspect, 4
 - purpose aspects, 48
 - query-focused, 64, 65
 - task, 24–26, 28–30, 33, 35–37
 - use, 5
- AutoSummENG, 39
- Basic Element, 4, 44
- Bayesian Model
 - crosslingual summarization, 33
 - query-focused summarization, 29
 - single document summarization, 26
- Centroid-based Approach
 - crosslingual summarization, 33
 - headline generation, 27
 - multi-document summarization, 28
- Clustering, 31
 - crosslingual summarization, 30, 32
 - multi-document summarization, 28, 29
 - multilingual summarization, 31, 32
- Coherence, 20, 91
- Cohesion, 19, 20, 26, 81, 83
 - document feature, 74
 - lexical chain, 68
 - speech summarization, 34
- Collocation
 - annotation, 59
- Conditional Random Fields
 - speech summarization, 34
- Content Feature
 - cue term, 27
 - length, 27
 - position, 27
- Coreference, 20
 - CLASSY, 28
 - inter-document, 28
 - resolution, 59
- Coreference Chain, 20, 69
 - single document summarization, 26
- Corpus
 - British National, 7

- CAST project, 7
- Reuters, 7
- Crosslingual Summarization, 23–25, 30–32, 35
 - DUC, 32
- Cue Term, 19, 22
 - crosslingual summarization, 32
 - headline generation, 27
 - LAKHAS, 32
- Decision Tree
 - headline generation, 27
 - single document summarization, 26
- Discourse Level Approach, 21, 22
 - document format, 21
 - rethorical structure, 22
 - thread of topics, 21
- Discourse Marker, 22
 - annotation, 59, 62, 63
 - lexicon, 63, 89
- Discourse Representation, 22
- Discourse Segment, 63
- Discursive Structure
 - Coherence, 21
 - Cohesion, 21
- Document structure, 3, 48
 - chapter, 3
 - header, 3
 - list, 3
 - section, 3
 - table, 3
- DUC
 - crosslingual summarization, 32
 - evaluation forum, 23
 - headline generation, 26
 - linguistic quality aspect, 30
 - manual evaluation, 38
 - multi-document summarization, 28
 - query-focused summarization, 29
 - responsiveness score, 30
 - ROUGE evaluation package, 38
 - ROUGE metric, 39
 - task, 22–24
 - meaning representation relation, 21
 - proximity, 19
 - similarity, 19
- EuroWordNet, 63
- Evaluation, 6
 - gold standard, 38
 - automatic, 38
 - AutoSummENG toolkit, 39
 - DUC, 23
 - extrinsic, 6
 - framework, 12
 - gold standard, 100
 - human, 38
 - intrinsic, 6, 7, 12, 100
 - linguistic quality aspect, 30
 - metric combination, 7
 - multi-document summarization, 38
 - NUGGETEER, 45
 - POURPRE, 45
 - precision metric, 7
 - pyramid method, 3, 40
 - QARLA framework, 7, 39
 - recall metric, 7
 - reference summary, 8
 - relative utility, 39
 - responsiveness score, 30
 - ROUGE metric, 7
 - ROUGE package, 7, 38
 - SEE, 38
 - single document summarization, 38
 - summary, 7
 - weight n-gram content, 39
- Evaluation Forum
 - CLEF, 76
 - DUC, 7, 17, 22–26, 29, 30, 35, 36
 - GALE, 45
 - MSE, 17, 32, 35, 36, 44
 - NTCIR, 6, 17
 - SUMMAC, 6, 17
 - TREC, 45, 76
- Evaluation Metric, 35
 - co-selection, 35
 - content-based, 35
 - discursive, 38
 - DUC, 38
- Entity Level Approach, 19
 - cohesion, 19
 - logical relation, 20

- grammatical, 38
- informational coverage, 38
- length requirement, 38
- responsiveness, 38
- Evaluator Agreement
 - Kappa measure, 7
- Evaluator Disagreement, 17
- Event-focused Summarization, 23
 - DUC, 23
- FEMsum, 47
 - algorithm
 - Antiredundancy SC, 91
 - Chunk Extractor SC, 90
 - Reordering SC, 92
 - Sentence Compression, 87
 - approaches, 54
 - LCsum, 55, 56, 59, 68
 - LEXsum, 56, 59, 68, 76, 77, 87
 - MLsum, 56, 59, 68
 - QAsum, 56, 64, 68, 76, 82, 87, 91
 - SEMsum, 56, 64, 68, 76, 77, 84, 87, 92
 - components, 50, 51, 54
 - Catalan LP, 62
 - Complex SC, 87
 - English LP, 63
 - Feature Extractor RID, 72
 - Generating QP, 65
 - Lexical Chain CE, 80
 - Lexical Chain RID, 68
 - Machine Learning CE, 80
 - Question & Answering CE, 82
 - Semantic CE, 83
 - Spanish LP, 62
 - Splitting QP, 65
 - instantiated component
 - Decision Trees CE, 80
 - Passage Retrieval RID, 76
 - qMDS Feature Extractor RID, 74
 - SDS Feature Extractor, 73
 - Support Vector Machine CE, 81
- FreeLing, 63
 - dependency parsing, 63
 - lemmatization, 63
 - morphological analysis, 63
 - NERC, 63
 - POS tagging, 63
 - shallow parsing, 63
 - tokenization, 63
- Graph-based Algorithm
 - crosslingual summarization, 33
 - MEAD, 28
 - multi-document summarization, 28
 - PageRank, 29, 33
 - query-focused summarization, 29
 - semantic, 81, 83
 - single document summarization, 26
- Headline Generation, 6, 11, 26, 27, 35
 - DUC, 23
- Hidden Markov Model
 - CLASSY, 28
 - crosslingual summarization, 33
 - headline generation, 27
 - multi-document summarization, 28
 - single document summarization, 26
 - speech summarization, 34
- Human Abstract, 15
- Inductive Logic Programing
 - single document summarization, 26
- Information Extraction, 22, 28
 - multi-document summarization, 28
 - single document summarization, 26
- Information Retrieval, 5, 20, 30, 76
 - JIRS PR, 77
 - lexical chain, 68
 - TALP-QA Passage Retrieval, 76
- Input
 - automatically translated, 24
 - creation scenario, 1, 17, 48
 - domain, 1, 3, 4, 17, 25, 34, 48
 - genre, 3, 4, 25, 48
 - ill-formed text, 4, 24, 30–36
 - language, 3, 4, 24, 30–32, 35, 37, 48
 - media, 1, 4, 24, 48
 - multi-document, 4
 - restricted domain, 3
 - single document, 4
 - specialization level, 3

- speech, 33, 34
- Just-the-news Summarization, 24, 166
- Keyphrase Extraction
 - multi-document summarization, 29
- Latent Semantic Analysis, 34
 - query-focused summarization, 30
 - single document summarization, 26
- Lemmatization, 63
- Length
 - headline generation, 27
- Lexical Chain, 20
 - representation, 68
 - single document summarization, 26
 - Word Sense Disambiguation, 69
- Lexical Similarity, 19, 20
 - crosslingual summarization, 30
 - distributional, 19
 - lexical chain, 20
 - orthographic, 19
 - thesaural relationship, 20
- Local Saliency, 20
- Logical Relation, 20
 - textual entailment, 21
- Logistic Regression
 - single document summarization, 26
- Machine Learning, 18, 80
 - crosslingual summarization, 33
 - decision tree, 80
 - feature extractor, 72
 - headline generation, 27
 - Mapping-Convergence algorithm, 82
 - multi-document summarization, 29, 68
 - multilingual summarization, 30
 - query-focused summarization, 29
 - single document summarization, 26, 68
 - speech summarization, 33
 - support vector machine, 81, 82
 - training corpus, 81, 162
- Machine Translation, 5, 24, 26, 30, 31, 35
 - crosslingual summarization, 30–32
 - LAKHAS, 32
 - multilingual summarization, 31
- Maximal Marginal Relevance, 34
 - multi-document summarization, 28
 - overcoming redundancy, 28
 - single document summarization, 26
- MEAD evaluation toolkit, 35
- MEAD project, 38
- Meaning Relation, 21
 - predicate-argument, 21
- Monolingual Summarization, 30
- Morphological Analysis, 63
- Multi-Document Summarization, 22–26, 28–30, 32, 35, 37, 48, 72
 - crosslingual, 30, 31
 - DUC, 23
 - feature extractor, 74
 - machine learning, 68
 - Passage Retrieval, 76, 77
 - query-focused, 29, 64
 - vs. single document summarization, 28
- Multilingual Summarization, 22, 30, 31, 35
 - machine translation, 31
 - MSE evaluation forum, 32, 44
- Multimedia Summarization, 4, 22
- Mutual Reinforcement Learning
 - multi-document summarization, 29
- Name Entity
 - Freeling NERC, 63
- Named Entity, 19, 32
 - ABIONET NERC, 63
- Natural Language Generation, 6
- Natural Language Processing, 3, 6, 15, 27
 - applications
 - Machine Translation, 102
 - techniques
 - similarity, 73, 74
- NIST Agencie, 7
 - assessor, 23
 - DUC, 23
- Noisy Channel Model
 - headline generation, 27
- Noun Phrase, 27
 - headline generation, 27
- Nugget
 - question & answering, 40

- Optical Character Recognizer, 4, 24
 - non-textual input, 4
 - textual output, 24
- Optimum Position Policy, 18
- Oracle Score
 - crosslingual summarization, 33
 - query-focused summarization, 29
- Passage Retrieval, 76
 - lexical chain, 69
- POS tagger, 63
- Position, 18, 22
 - crosslingual summarization, 32, 33
 - headline generation, 27
 - LAKHAS, 32
 - lexical chain, 80
 - multi-document summarization, 28
- Proximity, 19
- Pyramid Method, 40
 - annotation, 40
 - content selection, 41
 - gold standard, 40
 - informativeness score, 41
 - interannotator reliability, 40
 - linguistic quality, 41
 - metric, 40
 - modified pyramid score, 41
 - GISTexter, 30
 - multi-document summarization, 41
 - original pyramid score, 41
 - pyramid model, 40
 - summary content unit, 40
 - training corpus, 81
- QARLA
 - DUC, 39
 - JACK, 39
 - KING, 39
 - QUEEN, 39
- Query-focused Summarization, 23
 - DUC, 23, 24
 - real-world complex question, 24
- Question & Answering, 5, 24, 30
 - lexical chain, 69
 - QAsum, 56
 - query-focused summarization, 30
 - summary content extractor, 82
 - TALP-QA, 76, 82
- Random Indexing
 - single document summarization, 26
- Responsiveness
 - query-focused summarization, 30
- Rethorical parsing, 22
- Rethorical Relation, 22
- Rethorical Structure, 22
 - rhetorical argumentative compressor, 89
 - single document summarization, 26
- Rethorical Structure Theory, 22
- ROUGE
 - metric, 38, 39
 - multilingual summarization, 44
 - parametizable, 38
 - ROUGE-L, 38
 - ROUGE-n, 38
 - ROUGE-S_n, 38
 - ROUGE-SU_n, 38
 - ROUGE-W, 38
 - scoring function, 29
 - spontaneous speech summarization, 45
- Segmentation
 - discourse boundary, 3
 - ortographic, 3
 - syntactic, 3
 - topic, 3
- Segmenting Technique
 - lexical chain, 68
- Semantic Relations
 - query-focused summarization, 30
- Semantic Role Labeling, 59
- Semantic Tagging
 - euroWordNet, 59
 - WordNet, 59
- Sentence Reduction
 - CLASSY, 28
 - discourse boundary, 89
 - full parsing compressor, 87, 89
 - headline generation, 27
 - parse tree, 27
 - single document summarization, 26
- Similarity, 19

- centroid-based, 28
- crosslingual summarization, 30, 33
- document feature, 73
- FEMsum, 58
- graph representation, 20
- JIRS PR, 77
- lexical, 19
- lexical chain, 68
- lexico-semantic, 84
- maximal marginal relevance, 26
- multilingual summarization, 32
- multilingual unit, 31
- Named Entity, 91
- semantic, 19, 22, 81, 83
- sentence feature, 75
- vectorial distribution, 19
- Single Document Summarization, 17, 23–26, 35, 37, 48
 - crosslingual, 30
 - DUC, 23
 - feature extractor, 72
 - machine learning, 68
 - query-focused, 29
- Singular Value Decomposition, 39
- Smart Room, 123, 174
- Speech Summarization, 25, 33, 34, 55
 - broadcast news, 29, 34
 - corporate meeting, 34
 - lecture, 34
 - lexical chain, 72, 91
 - NERC, 63
 - political speech, 34
 - scientific presentation, 34
 - voice mail, 34
- Spontaneous Speech
 - evaluation, 45
- Statistical Model, 27
 - bayesian, 26, 27, 29
 - crosslingual summarization, 33
 - headline generation, 27
 - hidden markov model, 26
 - logistic regression, 26
 - multi-document summarization, 28
 - multilingual summarization, 30
 - single document summarization, 26
 - speech summarization, 34
- Summarization
 - abstract production, 15, 24, 25
 - crosslingual summarization, 36
 - embedded, 5
 - extract production, 24, 31
 - generic, 81
 - headline generation, 17, 36
 - monolingual, 24
 - multi-document, 17, 24, 35–37
 - multilingual, 17, 24, 31, 32, 37
 - multimedia, 17, 24
 - query-focused, 29, 36, 64, 65, 72, 74, 76, 77, 146, 147, 150, 155, 160, 161, 165, 166
 - single document, 24, 35–37
 - speech, 33, 34
 - spontaneous speech, 45
 - textual, 24
- Summarization phase, 15, 50, 51
 - analyzing, 16, 51
 - clarifying, 16, 51
 - extracting, 15, 31, 51
 - identifying relevant information, 15, 16, 33, 51
 - interpreting, 16, 31, 51
 - organizing, 15, 16, 51, 92
 - overcoming incoherence, 6, 15, 51
 - overcoming redundancy, 6, 15, 28, 51, 56, 81, 83, 85, 86, 91
 - post-editing, 15, 51
 - reducing, 15, 16, 27, 51
 - relevance ranking, 16, 29, 51
 - selecting, 15, 16, 29, 33, 51
 - stylizing, 16, 51
 - synthesizing, 15, 16, 51
 - topic identifying, 16, 31, 51
 - transforming, 16, 51
 - understanding, 16, 28, 51
- Summarizer
 - CENTRIFUSER, 19
 - English, 37
 - medical domain, 37
 - multi-document, 37
 - on-line, 37
 - CLASSY, 36
 - crosslingual, 33, 36

- multi-document, 28, 36
- query-focused, 36
- Copernic
 - download, 37
 - multilingual, 37
 - single document, 37
- DEMS, 3
- DEMS-MG, 36
 - crosslingual, 36
 - headline generation, 36
 - multi-document, 36
 - query-focused, 36
- DMSumm
 - download, 37
 - multilingual, 37
 - single document, 37
- Extractor
 - multilingual, 37
 - on-line, 37
 - single document, 37
- GISTexter, 36
 - crosslingual, 36
 - English, 37
 - headline generation, 36
 - multi-document, 36, 37
 - query-focused, 30, 36
 - single document, 37
- GistSumm
 - download, 37
 - multilingual, 37
 - single document, 37
- GLEANS, 36
 - crosslingual, 36
 - headline generation, 36
 - multi-document, 36
 - query-focused, 36
- Inxight Summarizer
 - multilingual, 37
 - single document, 37
- Island InText
 - English, 37
 - single document, 37
- LAKHAS
 - crosslingual, 32
- MEAD, 21
 - crosslingual, 33
 - headline generation, 27
 - multi-document, 28
- MEAD CLAIR
 - multi-document, 37
 - multilingual, 37
 - NewsInEssence, 37
 - on-line, 37
- MEAD Umichigan, 36
 - crosslingual, 36
 - headline generation, 36
 - multi-document, 36
 - query-focused, 36
- MS-Word Autosummarize
 - single document, 37
- MULTIGEN, 3
- multitask, 35
- NeATS ISI, 36
 - headline generation, 36
 - multi-document, 36
 - query-focused, 36
- Newsblaster
 - multi-document, 37
 - multilingual, 31, 37
- NTT, 19
- Pertinence
 - multilingual, 37
 - on-line, 37
 - single document, 37
- restricted domain, 24
- Sinope
 - download, 37
 - multilingual, 37
 - single document, 37
- Summ-It
 - on-line, 37
- SUMMAC, 20
- SUMMARIST, 18, 21
 - English, 31
 - Japanese, 31
 - multilingual, 31
 - Spanish, 31
- SumUM, 25, 36
 - crosslingual, 36
 - multi-document, 36

- query-focused, 36
- SweSum
 - multilingual, 37
 - on-line, 37
 - single document, 37
- TOPIARY
 - headline generation, 27
- Summary
 - abstract, 5
 - agregative, 5
 - automatic evaluation, 7, 8
 - background, 5, 50
 - biographies, 3
 - briefing, 5
 - content, 2, 50
 - critical, 5
 - definition, 1
 - event-focused content, 23
 - extract, 5
 - format, 5, 50
 - generic content, 5, 24, 35, 36
 - headline, 1
 - highlighting, 6
 - indicative, 5
 - informative, 5
 - just-the-news, 5, 50, 166
 - length, 2, 3, 6, 23, 50
 - manual evaluation, 7, 8, 12
 - media, 2
 - query-focused content, 5, 23–25, 29, 35, 36
 - sublanguage, 3
 - surrogate, 6
 - very short, 6
 - viewpoint-focused content, 23
- Summary Content Unit, 3, 40, 81
 - contributor, 41
- Summary Evaluation Environment, 38
- Summary Example
 - book synopsis, 1
 - film synopsis, 1
 - meeting minute, 1
 - stock market bulletin, 1
 - student outline, 1
 - weather forecast, 1
- Support Vector Machine, 81
 - crosslingual summarization, 33
 - multi-document summarization, 29
- Surface Level Approach, 18, 19
 - bias, 18
 - cue term, 19
 - location, 18
 - term frequency, 18
- Syntactic Analysis, 19, 59
 - Collin's parser, 64
 - dependency parsing, 63, 64
 - document feature, 74
 - full parsing compressor, 89
 - headline generation, 27
 - MINIPAR, 64
 - query-focused summarization, 30
 - sentence reduction, 87
 - shallow parsing, 63
 - sint representation, 59, 61, 62, 66, 76, 82
 - summary composer, 56
- Term Frequency, 18, 22, 39
 - crosslingual summarization, 32
 - headline generation, 27
 - LAKHAS, 32
 - lexical chain, 72
 - query-focused summarization, 29
 - spontaneous speech, 48
 - TFIDF feature, 75
- Term frequency
 - document feature, 75
- Term Frequency Inverse Document Frequency, 39
- Text Summarization, 17, 26
 - subproblems, 16
- Textual Entailment, 21
 - GISTexter, 30
 - multi-document summarization, 30
 - query-focused summarization, 30
- Textual Unit, 3, 32, 33
 - Basic Element, 44
 - basic element, 4
 - clause, 3, 6, 33, 87
 - discourse boundary, 63, 90
 - discourse segment, 63
 - elementary discourse unit, 39
 - factoid, 40, 45

- meaning unit, 3
- multilingual unit, 31
- multiparagraph, 6
- nugget, 40, 45
- paragraph, 3, 6, 100
- phrase, 6
- phrase chunk, 33, 64, 87, 90
- segmenting technique, 59
- segmenting techniques, 52
- semantic dialogue unit, 33
- sentence, 3, 32, 33, 62, 65, 66, 73, 75, 77, 78, 81, 87, 91, 92
- summary content unit, 3, 40, 45
- topic passage, 6
- utterance, 34
- Tokenization, 63
- Topic Identification, 21
 - headline generation, 27
 - lexical chain, 68
 - MEAD, 21
 - single document summarization, 26
 - SUMMARIST, 21
 - topic signature, 21
- Translation
 - source to summary language, 27
- Tree Similarity, 30
- Trigger Word, 27
- Unigram Language Model
 - headline generation, 27
- Unsupervised Topic Discovery
 - headline generation, 27
- User Need
 - generic content, 36
 - natural language query, 29, 30, 35
 - query-focused content, 36
- Vector Space Model
 - query-focused summarization, 30
- Viewpoint-focused Summarization, 23
 - DUC, 23
- Weighted Bipartite Graph
 - multi-document summarization, 29
- Word Sense Disambiguation, 85
 - euroWordNet, 63
 - lexical chain, 68, 72
- World Wide Web
 - resource, 28, 178
 - searcher, 24, 29