keyword_analysis

November 9, 2017

In [15]:
```
import pickle
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [16]:
```
dat = pickle.load(open("extracted_files/extracted_raw.p", "rb"))
dat = list(filter(None, dat))
df = pd.DataFrame(dat)
```

In [17]:
```
df.columns
```

Out[17]:
```
Index(['abstract', 'authors', 'cite_count', 'cover_date', 'doi', 'keywords',
       'publication_name', 'reference_count', 'subject_area', 'title', 'type',
       'volume'],
      dtype='object')
```

In [18]:
```
abstracts = " ".join(list(df.abstract))
```

In [19]:
```
abstracts = abstracts.lower()
```

In [20]:
```
from nltk.tokenize import sent_tokenize, word_tokenize
```

In [21]:
```
words = word_tokenize(abstracts)
```

In [22]:
```
words
```

Out[22]:
```
['ï',
 '2017',
 'elsevier',
 'ltd',
 'a',
 'comprehensive',
 'public',
 'health',
 'strategy',
 'for',
 'adolescent',
```

```
'suicide',
'prevention',
'includes',
'upstream',
'prevention',
'strategies',
',',
'strategies',
'for',
'risk',
'recognition',
',',
'and',
'services',
'for',
'those',
'at',
'risk',
'.',
'interpersonal',
'trauma',
'and',
'substance',
'use',
'are',
'important',
'prevention',
'targets',
'as',
'each',
'is',
'associated',
'with',
'risk',
'for',
'suicide',
'attempts',
'.',
'multiple',
'prevention',
'programs',
'target',
'these',
'factors',
';',
'however',
',',
'the',
```

```
'family',
'check-up',
',',
'designed',
'to',
'reduce',
'substance',
'use',
'and',
'behavioral',
'problems',
',',
'also',
'has',
'been',
'associated',
'with',
'reduced',
'suicide',
'risk',
'.',
'several',
'youth',
'screening',
'instruments',
'have',
'shown',
'utility',
',',
'and',
'a',
'large-scale',
'trial',
'is',
'underway',
'to',
'develop',
'a',
'computerized',
'adaptive',
'screen',
'.',
'similarly',
',',
'several',
'types',
'of',
'psychotherapy',
```

```
'have',
'shown',
'promise',
',',
'and',
'sufficiently',
'powered',
'studies',
'are',
'underway',
'to',
'provide',
'more',
'definitive',
'results',
'.',
'the',
'climbing',
'youth',
'suicide',
'rate',
'warrants',
'an',
'urgent',
',',
'concerted',
'effort',
'to',
'develop',
'and',
'implement',
'effective',
'prevention',
'strategies',
'.',
'ī',
'2017',
'elsevier',
'b.v.',
'we',
'live',
'in',
'a',
'world',
'were',
'data',
'are',
'generated',
```

```
'from',
'a',
'myriad',
'of',
'sources',
',',
'and',
'it',
'is',
'really',
'cheap',
'to',
'collect',
'and',
'storage',
'such',
'data',
'.',
'however',
',',
'the',
'real',
'benefit',
'is',
'not',
'related',
'to',
'the',
'data',
'itself',
',',
'but',
'with',
'the',
'algorithms',
'that',
'are',
'capable',
'of',
'processing',
'such',
'data',
'in',
'a',
'tolerable',
'elapse',
'time',
',',
```

```
'and',
'to',
'extract',
'valuable',
'knowledge',
'from',
'it',
'.',
'therefore',
',',
'the',
'use',
'of',
'big',
'data',
'analytics',
'tools',
'provide',
'very',
'significant',
'advantages',
'to',
'both',
'industry',
'and',
'academia',
'.',
'the',
'mapreduce',
'programming',
'framework',
'can',
'be',
'stressed',
'as',
'the',
'main',
'paradigm',
'related',
'with',
'such',
'tools',
'.',
'it',
'is',
'mainly',
'identified',
'by',
```

```
'carrying',
'out',
'a',
'distributed',
'execution',
'for',
'the',
'sake',
'of',
'providing',
'a',
'high',
'degree',
'of',
'scalability',
',',
'together',
'with',
'a',
'fault-tolerant',
'scheme',
'.',
'in',
'every',
'mapreduce',
'algorithm',
',',
'first',
'local',
'models',
'are',
'learned',
'with',
'a',
'subset',
'of',
'the',
'original',
'data',
'within',
'the',
'so-called',
'map',
'tasks',
'.',
'then',
',',
'the',
```

```
'reduce',
'task',
'is',
'devoted',
'to',
'fuse',
'the',
'partial',
'outputs',
'generated',
'by',
'each',
'map',
'.',
'the',
'ways',
'of',
'designing',
'such',
'fusion',
'of',
'information/models',
'may',
'have',
'a',
'strong',
'impact',
'in',
'the',
'quality',
'of',
'the',
'final',
'system',
'.',
'in',
'this',
'work',
',',
'we',
'will',
'enumerate',
'and',
'analyze',
'two',
'alternative',
'methodologies',
'that',
```

```
'may',
'be',
'found',
'both',
'in',
'the',
'specialized',
'literature',
'and',
'in',
'standard',
'machine',
'learning',
'libraries',
'for',
'big',
'data',
'.',
'our',
'main',
'objective',
'is',
'to',
'provide',
'an',
'introduction',
'of',
'the',
'characteristics',
'of',
'these',
'methodologies',
',',
'as',
'well',
'as',
'giving',
'some',
'guidelines',
'for',
'the',
'design',
'of',
'novel',
'algorithms',
'in',
'this',
'field',
```

```
'of',
'research',
'.',
'finally',
',',
'a',
'short',
'experimental',
'study',
'will',
'allow',
'us',
'to',
'contrast',
'the',
'scalability',
'issues',
'for',
'each',
'type',
'of',
'process',
'fusion',
'in',
'mapreduce',
'for',
'big',
'data',
'analytics',
'.',
'I',
'2017',
'the',
'naïve',
'bayes',
'switching',
'linear',
'dynamical',
'system',
'(',
'nb-slds',
')',
'is',
'proposed',
'as',
'a',
'novel',
'variant',
```

```
'of',
'the',
'switching',
'linear',
'dynamical',
'system',
'(',
'slds',
')',
'.',
'the',
'variant',
'models',
'multi-variable',
'systems',
'that',
'undergo',
'regime',
'changes',
'in',
'their',
'dynamics',
'.',
'the',
'model',
'may',
'be',
'applied',
'to',
'identify',
'regime',
'changes',
'or',
'classify',
'systems',
'according',
'to',
'their',
'dynamics',
'.',
'the',
'nb-slds',
'provides',
'the',
'means',
'to',
'fuse',
'multiple',
```

```
'sequential',
'data',
'sources',
'into',
'a',
'single',
'model',
'.',
'a',
'key',
'feature',
'of',
'the',
'model',
'is',
'that',
'it',
'is',
'able',
'to',
'handle',
'missing',
'and',
'unsynchronised',
'data',
'.',
'filtering',
'and',
'smoothing',
'algorithms',
'for',
'inference',
'and',
'an',
'expectation',
'maximisation',
'algorithm',
'for',
'parameter',
'learning',
'in',
'the',
'nb-slds',
'are',
'presented',
'.',
'the',
'model',
```

```
'is',
'demonstrated',
'and',
'compared',
'to',
'the',
'slds',
'and',
'hidden',
'markov',
'model',
'(',
'hmm',
')',
'in',
'a',
'human',
'action',
'recognition',
'problem',
'.',
'Ī',
'2017',
'elsevier',
'b.v.',
'present',
'measures',
'of',
'the',
'degree',
'of',
'agreement',
'in',
'group',
'decision-making',
'using',
'hesitant',
'fuzzy',
'linguistic',
'term',
'sets',
'allow',
'consensus',
'or',
'agreement',
'measurement',
'when',
'decision',
```

```
'makers',
'',
'assessments',
'involve',
'hesitance',
'.',
'yet',
'they',
'do',
'not',
'discriminate',
'with',
'different',
'degrees',
'of',
'consensus',
'among',
'situations',
'with',
'discordant',
'or',
'polarized',
'assessments',
'.',
'the',
'visualization',
'of',
'differences',
'among',
'groups',
'for',
'which',
'there',
'is',
'no',
'agreement',
'but',
'different',
'possible',
'levels',
'of',
'disagreement',
'is',
'an',
'important',
'issue',
'in',
'collective',
```

```
'decision-making',
'situations',
'.',
'in',
'this',
'paper',
',',
'we',
'propose',
'new',
'collective',
'and',
'individual',
'consensus',
'measures',
'that',
'explicitly',
'consider',
'the',
'hesitance',
'of',
'the',
'decision',
'makers',
'',
'hesitance',
'in',
'giving',
'an',
'opinion',
'and',
'also',
'the',
'gap',
'between',
'non-overlapping',
'assessments',
',',
'thus',
'allowing',
'the',
'measurement',
'of',
'the',
'polarization',
'present',
'within',
'the',
```

```
'group',
"'s",
'opinions',
'.',
'in',
'addition',
',',
'an',
'expert',
"'s",
'profile',
'is',
'defined',
'by',
'considering',
'the',
'expert',
"'s",
'behavior',
'in',
'previous',
'assessments',
'in',
'group',
'decision-making',
'processes',
'in',
'terms',
'of',
'precision',
'and',
'dissension',
'.',
'©',
'2017',
'elsevier',
'b.v.',
'biometric',
'verification',
'systems',
'are',
'currently',
'being',
'deployed',
'in',
'numerous',
'large-scale',
'and',
```

```
'everyday',
'applications',
'.',
'it',
'is',
'hence',
'of',
'the',
'utmost',
'importance',
'to',
'protect',
'the',
'privacy',
'of',
'the',
'enrolled',
'subjects',
'.',
'biometric',
'template',
'protection',
'schemes',
'are',
'designed',
'to',
'protect',
'biometric',
'reference',
'data',
'in',
'an',
'irreversible',
'and',
'unlinkable',
'manner',
',',
'while',
'maintaining',
'key',
'system',
'properties',
'like',
'the',
'accuracy',
'or',
'the',
'speed',
```

```
'.',
'in',
'past',
'years',
',',
'template',
'protection',
'schemes',
'based',
'on',
'bloom',
'filters',
'have',
'been',
'introduced',
'and',
'applied',
'to',
'various',
'biometric',
'characteristics',
'.',
'while',
'the',
'irreversibility',
'and',
'unlinkability',
'of',
'bloom',
'filter-based',
'protection',
'schemes',
'have',
'been',
'shown',
',',
'their',
'application',
'to',
'any',
'given',
'unprotected',
'template',
'is',
'not',
'straightforward',
'.',
'in',
```

```
'this',
'article',
'we',
'present',
'a',
'methodology',
'for',
'the',
'estimation',
'of',
'the',
'main',
'parameters',
'of',
'such',
'schemes',
',',
'based',
'on',
'a',
'statistical',
'analysis',
'of',
'the',
'unprotected',
'templates',
'.',
'furthermore',
',',
'in',
'order',
'to',
'increase',
'verification',
'accuracy',
'and',
'privacy',
'protection',
',',
'a',
'general',
'approach',
'for',
'a',
'protected',
'weighted',
'feature',
'level',
```

```
'fusion',
'is',
'proposed',
'.',
'in',
'order',
'to',
'avoid',
'biased',
'results',
',',
'the',
'soundness',
'of',
'the',
'estimation',
'methodologies',
'is',
'confirmed',
'for',
'face',
',',
'iris',
',',
'fingerprint',
'and',
'fingervein',
'over',
'two',
'totally',
'different',
'sets',
'of',
'publicly',
'available',
'databases',
'.',
'in',
'addition',
',',
'we',
'show',
'how',
'the',
'weighted',
'feature',
'level',
'fusion',
```

```
'preserves',
'the',
'accuracy',
'of',
'the',
'unprotected',
'score',
'level',
'fusion',
',',
'while',
'it',
'adds',
'privacy',
'protection',
'to',
'the',
'system',
'.',
'Ï',
'2017',
'elsevier',
'b.v.',
'the',
'separation',
'of',
'malignant',
'from',
'benign',
'lung',
'nodules',
'on',
'chest',
'computed',
'tomography',
'(',
'ct',
')',
'is',
'important',
'for',
'the',
'early',
'detection',
'of',
'lung',
'cancer',
',',
```

```
 'since',
 'early',
 'detection',
 'and',
 'management',
 'offer',
 'the',
 'best',
 'chance',
 'for',
 'cure',
 '.',
 'although',
 'deep',
 'learning',
 'methods',
 'have',
 'recently',
 'produced',
 'a',
 'marked',
 'improvement',
 'in',
 'image',
 'classification',
 'there',
 'are',
 'still',
 'challenges',
 ...]
```

```
In [23]: keywords = pickle.load(open("keywords.p", "rb"))
```

```
In [24]: also_keywords = []
         for key in keywords:
             also_keywords.append(key.split())

         keywords = []
         for sublist in also_keywords:
             for item in sublist:
                 keywords.append(item.lower())
         keywords
```

```
Out[24]: ['abductive',
          'logic',
          'programming',
          'abductive',
          'reasoning',
```

```
'abstract',
'data',
'type',
'abstraction',
'accelerating',
'change',
'action',
'language',
'action',
'model',
'learning',
'action',
'selection',
'adaptive',
'algorithm',
'adaptive',
'neuro',
'fuzzy',
'inference',
'system',
'admissible',
'heuristic',
'affective',
'computing',
'agent',
'architecture',
'ai',
'accelerator',
'ai-complete',
'algorithm',
'algorithmic',
'efficiency',
'algorithmic',
'probability',
'alphago',
'ambient',
'intelligence',
'analysis',
'of',
'algorithms',
'answer',
'set',
'programming',
'anytime',
'algorithm',
'application',
'programming',
'interface',
```

```
'approximate',
'string',
'matching',
'approximation',
'error',
'argumentation',
'framework',
'artificial',
'immune',
'system',
'artificial',
'intelligence',
'artificial',
'intelligence',
'markup',
'language',
'artificial',
'neural',
'network',
'association',
'for',
'the',
'advancement',
'of',
'artificial',
'intelligence',
'asymptotic',
'computational',
'complexity',
'attributional',
'calculus',
'augmented',
'reality',
'automata',
'theory',
'automated',
'planning',
'and',
'scheduling',
'automated',
'reasoning',
'autonomic',
'computing',
'autonomous',
'car',
'autonomous',
'robot',
'backpropagation',
```

```
'backward',
'chaining',
'batch',
'normalisation',
'bayesian',
'programming',
'bees',
'algorithm',
'behavior',
'informatics',
'behavior',
'tree',
'belief-desire-intention',
'software',
'model',
'bias,variance',
'tradeoff',
'big',
'data',
'big',
'o',
'notation',
'binary',
'tree',
'bio-inspired',
'computing',
'blackboard',
'system',
'boolean',
'satisfiability',
'problem',
'brain',
'technology',
'branching',
'factor',
'brute-force',
'search',
'case-based',
'reasoning',
'chatterbot',
'cloud',
'robotics',
'cluster',
'analysis',
'cobweb',
'cognitive',
'architecture',
'cognitive',
```

```
'computing',
'cognitive',
'science',
'combinatorial',
'optimization',
'committee',
'machine',
'commonsense',
'knowledge',
'commonsense',
'reasoning',
'computational',
'chemistry',
'computational',
'complexity',
'theory',
'computational',
'creativity',
'computational',
'humor',
'computational',
'intelligence',
'computational',
'learning',
'theory',
'computational',
'linguistics',
'computational',
'mathematics',
'computational',
'neuroscience',
'computational',
'number',
'theory,computational',
'problem',
'computational',
'statistics',
'computational',
'vision',
'computer-automated',
'design',
'computer',
'science',
'computer',
'vision',
'connectionism',
'consistent',
'heuristic',
```

```
'constraint',
'logic',
'programming',
'constraint',
'programming',
'constructed',
'language',
'control',
'theory',
'convolutional',
'neural',
'network',
'crossover',
'darkforest',
'dartmouth',
'workshop',
'data',
'fusion',
'data',
'integration',
'data',
'mining',
'data',
'science',
'data',
'set',
'data',
'warehouse',
'datalog',
'decision',
'boundary',
'decision',
'support',
'system',
'decision',
'theory',
'declarative',
'programming',
'deductive',
'classifier',
'deep',
'blue',
'deep',
'learning',
'default',
'logic',
'description',
'logic',
```

```
'developmental',
'robotics',
'diagnosis',
'dialog',
'system',
'dimensionality',
'reduction',
'discrete',
'system',
'distributed',
'artificial',
'intelligence',
'dynamic',
'epistemic',
'logic',
'embodied',
'agent',
'embodied',
'cognitive',
'science',
'ensemble',
'averaging',
'ethics',
'of',
'artificial',
'intelligence',
'evolutionary',
'algorithm',
'evolutionary',
'computation',
'evolving',
'classification',
'function',
'existential',
'risk',
'expert',
'systems',
'feature',
'extraction',
'feature',
'selection',
'first-order',
'logic',
'fluent',
'formal',
'language',
'forward',
'chaining',
```

```
'frame',
'frame',
'language',
'frame',
'problem',
'friendly',
'artificial',
'intelligence',
'futures',
'studies',
'fuzzy',
'control',
'system',
'fuzzy',
'logic',
'fuzzy',
'set',
'game',
'theory',
'genetic',
'algorithm',
'genetic',
'operator',
'glowworm',
'swarm',
'optimization',
'google',
'deepmind',
'graph',
'graph',
'graph',
'database',
'graph',
'theory',
'graph',
'traversal',
'heuristic',
'hidden',
'layerhidden',
'unitieee',
'computational',
'intelligence',
'society',
'incremental',
'learning',
'inference',
'engine',
'information',
```

```
'integration',
'information',
'processing',
'language',
'intelligence',
'amplification',
'intelligence',
'explosion',
'intelligent',
'agent',
'intelligent',
'control',
'intelligent',
'personal',
'assistant',
'interpretation',
'issue',
'trees',
'kernel',
'method',
'kl-one',
'knowledge',
'acquisition',
'knowledge-based',
'systems',
'knowledge',
'engineering',
'knowledge',
'extraction',
'knowledge',
'interchange',
'format',
'knowledge',
'representation',
'and',
'reasoning',
'linked',
'data',
'lisp',
'logic',
'programming',
'machine',
'vision',
'markov',
'chain',
'markov',
'decision',
'process',
```

```
'mathematical',
'optimization',
'machine',
'learning',
'machine',
'listening',
'machine',
'perception',
'mechanism',
'design',
'mechatronics',
'metabolic',
'network',
'modelling',
'metaheuristic',
'model',
'checking',
'modus',
'ponens',
'modus',
'tollens',
'monte',
'carlo',
'tree',
'search',
'multi-agent',
'system',
'multi-swarm',
'optimization',
'mutation',
'mycin',
'name',
'binding',
'named-entity',
'recognition',
'named',
'graph',
'natural',
'language',
'processing',
'natural',
'language',
'programming',
'network',
'motif',
'neuro-fuzzy',
'neurocybernetics',
'node',
```

```
'nondeterministic',
'algorithm',
'nouvelle',
'ai',
'np',
'np-completeness',
'np-hardness',
'offline',
'learning',
'online',
'learning',
'ontology',
'engineering',
'ontology',
'learning',
'openai',
'opencog',
'open',
'mind',
'common',
'sense',
'open-source',
'software',
'partial',
'order',
'reduction',
'partially',
'observable',
'markov',
'decision',
'process',
'particle',
'swarm',
'optimization',
'pathfinding',
'pattern',
'recognition',
'planner',
'predicate',
'logic',
'principal',
'component',
'analysis',
'principle',
'of',
'rationality',
'probabilistic',
'programming',
```

```
'language',
'production',
'rule',
'representation',
'production',
'system',
'programming',
'language',
'prolog',
'propositional',
'calculus',
'python',
'qualification',
'problem',
'quantifier',
'query',
'language',
'r',
'programming',
'language',
'reasoning',
'system',
'recurrent',
'neural',
'network',
'region',
'connection',
'calculus',
'reinforcement',
'learning',
'resource',
'description',
'framework',
'rete',
'algorithm',
'robotics',
'rule-based',
'system',
'satisfiability',
'search',
'algorithm',
'selection',
'self-management',
'semantic',
'network',
'semantic',
'reasoner',
'semantic',
```

```
'query',
'semantics',
'sensor',
'fusion',
'separation',
'logic',
'similarity',
'learning',
'simulated',
'annealing',
'situated',
'approach',
'situation',
'calculus',
'sld',
'resolution',
'soft',
'computing',
'software',
'software',
'engineering',
'spatial-temporal',
'reasoning',
'sparql',
'speech',
'recognition',
'spiking',
'neural',
'network',
'state',
'statistical',
'classification',
'statistical',
'relational',
'learning',
'stochastic',
'optimization',
'stochastic',
'semantic',
'analysisstrips',
'subject-matter',
'expert',
'superintelligence',
'supervised',
'learning',
'swarm',
'intelligence',
'symbolic',
```

```
'artificial',
'intelligence',
'synthetic',
'intelligence',
'systems',
'neuroscience',
'technological',
'singularity',
'temporal',
'difference',
'learning',
'tensor',
'network',
'theory',
'tensorflow',
'theoretical',
'computer',
'science',
'theory',
'of',
'computation',
'thompson',
'sampling',
'time',
'complexity',
'transhumanism',
'transition',
'system',
'tree',
'traversal',
'true',
'quantified',
'boolean',
'formula',
'turing',
'test',
'type',
'system',
'unsupervised',
'learning',
'vision',
'processing',
'unit',
'watson',
'weak',
'ai',
'world',
'wide',
```

```
          'web',
          'consortium']

In [25]: key_dict = {}
         for key in keywords:
             for word in words:
                 if word.lower() == key.lower():
                     if key in key_dict.keys():
                         key_dict[key] += 1
                     else:
                         key_dict[key] = 1

In [26]: key_dict

Out[26]: {'abstract': 12,
          'abstraction': 4,
          'accelerating': 2,
          'accelerator': 2,
          'acquisition': 22,
          'action': 84,
          'adaptive': 138,
          'admissible': 2,
          'advancement': 7,
          'agent': 99,
          'ai': 117,
          'algorithm': 5418,
          'algorithmic': 22,
          'algorithms': 366,
          'alphago': 3,
          'ambient': 16,
          'amplification': 1,
          'analysis': 1107,
          'and': 11298,
          'annealing': 7,
          'answer': 8,
          'anytime': 1,
          'application': 149,
          'approach': 418,
          'approximate': 13,
          'approximation': 20,
          'architecture': 82,
          'argumentation': 11,
          'artificial': 2754,
          'assistant': 4,
          'association': 37,
          'asymptotic': 2,
          'augmented': 5,
          'automata': 9,
```

```
'automated': 54,
'autonomic': 1,
'autonomous': 44,
'averaging': 5,
'backpropagation': 3,
'backward': 1,
'batch': 9,
'bayesian': 32,
'bees': 3,
'behavior': 158,
'big': 104,
'binary': 36,
'binding': 1,
'bio-inspired': 7,
'blue': 8,
'boolean': 16,
'boundary': 19,
'brain': 29,
'branching': 6,
'brute-force': 1,
'calculus': 12,
'car': 5,
'carlo': 6,
'case-based': 7,
'chain': 23,
'chaining': 2,
'change': 32,
'checking': 17,
'chemistry': 1,
'classification': 504,
'classifier': 64,
'cloud': 62,
'cluster': 37,
'cognitive': 168,
'combinatorial': 19,
'common': 54,
'complexity': 210,
'component': 38,
'computation': 90,
'computational': 1988,
'computer': 258,
'computing': 390,
'connection': 12,
'consistent': 12,
'constraint': 78,
'constructed': 54,
'control': 423,
'convolutional': 15,
```

```
'creativity': 1,
'crossover': 6,
'data': 8352,
'database': 57,
'decision': 830,
'deep': 134,
'default': 2,
'description': 52,
'design': 344,
'diagnosis': 79,
'dialog': 1,
'difference': 30,
'dimensionality': 25,
'discrete': 47,
'distributed': 56,
'dynamic': 102,
'efficiency': 109,
'engine': 13,
'engineering': 144,
'ensemble': 47,
'epistemic': 1,
'error': 78,
'evolutionary': 94,
'evolving': 10,
'existential': 3,
'expert': 112,
'extraction': 106,
'factor': 24,
'feature': 374,
'first-order': 4,
'for': 2084,
'formal': 27,
'format': 6,
'formula': 21,
'forward': 24,
'frame': 33,
'framework': 268,
'function': 108,
'fusion': 150,
'fuzzy': 720,
'game': 51,
'genetic': 142,
'google': 8,
'graph': 474,
'heuristic': 102,
'hidden': 25,
'immune': 5,
'incremental': 12,
```

```
'inference': 78,
'informatics': 3,
'information': 786,
'integration': 76,
'intelligence': 2590,
'intelligent': 303,
'interchange': 2,
'interface': 46,
'interpretation': 14,
'issue': 48,
'kernel': 22,
'knowledge': 948,
'knowledge-based': 5,
'language': 528,
'learning': 6258,
'linguistics': 1,
'linked': 9,
'logic': 550,
'machine': 1150,
'markov': 54,
'matching': 36,
'mathematical': 34,
'mathematics': 6,
'mechanism': 66,
'metabolic': 2,
'metaheuristic': 23,
'method': 643,
'mind': 10,
'mining': 68,
'model': 2052,
'modelling': 42,
'modus': 4,
'monte': 7,
'multi-agent': 12,
'mutation': 14,
'name': 2,
'named': 25,
'natural': 120,
'network': 2888,
'neural': 932,
'neuro-fuzzy': 10,
'neuroscience': 2,
'node': 31,
'nondeterministic': 1,
'normalisation': 2,
'notation': 3,
'number': 219,
'o': 18,
```

```
'observable': 3,
'of': 38465,
'offline': 7,
'online': 69,
'ontology': 38,
'open': 33,
'open-source': 5,
'opencog': 1,
'operator': 24,
'optimization': 1692,
'order': 214,
'partial': 41,
'partially': 8,
'particle': 54,
'pattern': 62,
'perception': 13,
'personal': 30,
'planner': 4,
'planning': 37,
'ponens': 1,
'predicate': 4,
'principal': 16,
'principle': 13,
'probabilistic': 28,
'probability': 35,
'problem': 1668,
'process': 570,
'processing': 351,
'production': 96,
'programming': 744,
'prolog': 1,
'propositional': 2,
'python': 2,
'qualification': 4,
'quantified': 9,
'query': 32,
'r': 13,
'rationality': 1,
'reality': 10,
'reasoner': 1,
'reasoning': 280,
'recognition': 411,
'recurrent': 17,
'reduction': 100,
'region': 37,
'reinforcement': 16,
'relational': 16,
'representation': 170,
```

```
'resolution': 19,
'resource': 33,
'risk': 33,
'robot': 39,
'robotics': 36,
'rule': 79,
'rule-based': 17,
'sampling': 13,
'satisfiability': 12,
'scheduling': 52,
'science': 255,
'search': 360,
'selection': 360,
'semantic': 204,
'semantics': 10,
'sense': 22,
'sensor': 50,
'separation': 5,
'set': 669,
'similarity': 62,
'simulated': 26,
'singularity': 3,
'situated': 4,
'situation': 22,
'society': 15,
'soft': 16,
'software': 360,
'sparql': 2,
'spatial-temporal': 3,
'speech': 21,
'spiking': 1,
'state': 67,
'statistical': 148,
'statistics': 15,
'stochastic': 54,
'string': 6,
'studies': 100,
'supervised': 30,
'support': 166,
'swarm': 282,
'symbolic': 8,
'synthetic': 31,
'system': 8138,
'systems': 990,
'technological': 17,
'technology': 115,
'temporal': 40,
'tensor': 8,
```

```
 'test': 86,
 'the': 12384,
 'theoretical': 37,
 'theory': 774,
 'time': 304,
 'tollens': 1,
 'tradeoff': 5,
 'transition': 13,
 'traversal': 2,
 'tree': 172,
 'trees': 14,
 'true': 19,
 'turing': 3,
 'type': 116,
 'unit': 31,
 'unsupervised': 23,
 'vision': 184,
 'warehouse': 9,
 'weak': 10,
 'web': 34,
 'wide': 26,
 'world': 51}
```

In [27]:
```
import operator
# x = {1: 2, 3: 4, 4: 3, 2: 1, 0: 0}
sorted_x = sorted(key_dict.items(), key=operator.itemgetter(1))
```

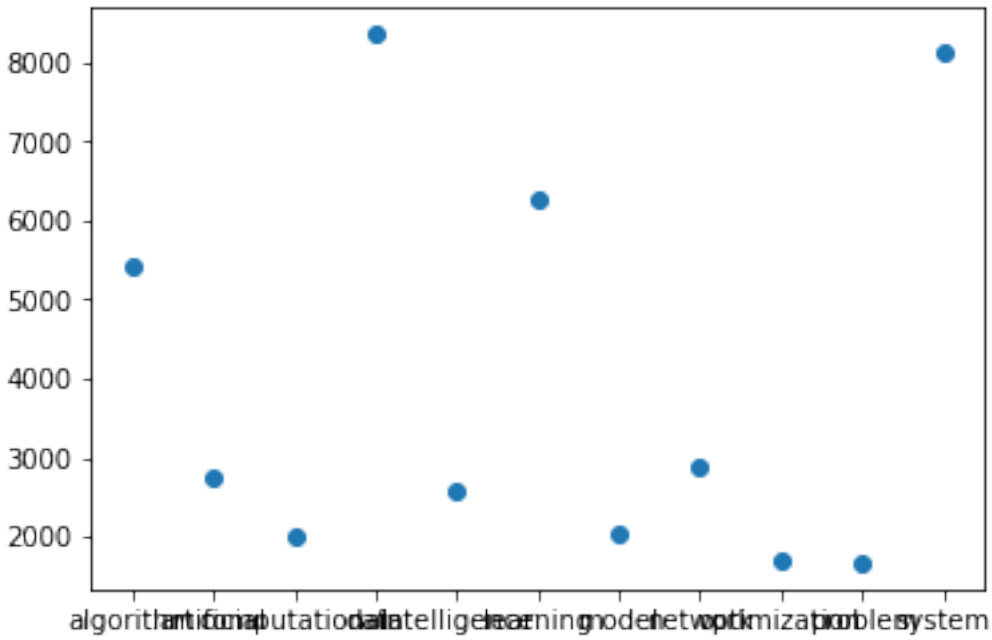
In [28]:
```
for_plotting = sorted_x[-15:]
```

In [29]:
```
for_plotting
```

Out[29]:
```
[('problem', 1668),
 ('optimization', 1692),
 ('computational', 1988),
 ('model', 2052),
 ('for', 2084),
 ('intelligence', 2590),
 ('artificial', 2754),
 ('network', 2888),
 ('algorithm', 5418),
 ('learning', 6258),
 ('system', 8138),
 ('data', 8352),
 ('and', 11298),
 ('the', 12384),
 ('of', 38465)]
```

In [30]:
```
to_remove = "of,the,and,for,a,i,in,an"
to_remove = to_remove.split(",")
to_remove
```

```
Out[30]: ['of', 'the', 'and', 'for', 'a', 'i', 'in', 'an']

In [31]: # for i, val in enumerate(for_plotting):
         #     print(i,val)
         #     if key in to_remove:
         i = 0
         while 1:
             if for_plotting[i][0] in to_remove:
                 del for_plotting[i]
             else:
                 i += 1
             if i >= len(for_plotting):
                 break

In [32]: for_plotting

Out[32]: [('problem', 1668),
          ('optimization', 1692),
          ('computational', 1988),
          ('model', 2052),
          ('intelligence', 2590),
          ('artificial', 2754),
          ('network', 2888),
          ('algorithm', 5418),
          ('learning', 6258),
          ('system', 8138),
          ('data', 8352)]

In [33]: x = []
         y = []
         for val in for_plotting:
             x.append(val[0])
             y.append(val[1])
         # pickle.dump([x,y], open("for_plotting.p", "wb"))

In [34]: # x, y = pickle.load(open("for_plotting.p", "rb"))

In [35]: plt.scatter(x,y)
         plt.show()
```




```
In [37]: li = list(df.subject_area)

In [38]: flat_li = []
         for sublist in li:
             for item in sublist:
                 flat_li.append(item)

In [39]: flat_li

Out[39]: ['Psychology (all)',
          'Software',
          'Signal Processing',
          'Information Systems',
          'Hardware and Architecture',
          'Software',
          'Signal Processing',
          'Information Systems',
          'Hardware and Architecture',
          'Software',
          'Signal Processing',
          'Information Systems',
          'Hardware and Architecture',
          'Software',
          'Signal Processing',
          'Information Systems',
          'Hardware and Architecture',
```

```
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Decision Sciences (all)',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Computer Science (all)',
'Computer Science (all)',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
```

```
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Computer Science (all)',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Analysis',
'Discrete Mathematics and Combinatorics',
'Applied Mathematics',
'Analytical Chemistry',
'Food Science',
'Numerical Analysis',
'Modeling and Simulation',
'Applied Mathematics',
'Neuroscience (all)',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
```

```
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Soil Science',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science (all)',
'Computer Science (all)',
'Computer Science (all)',
'Computer Science (all)',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Computer Science Applications',
```

```
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science (all)',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computer Science (all)',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Algebra and Number Theory',
'Computational Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
```

```
'Applied Mathematics',
'Algebra and Number Theory',
'Computational Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Industrial and Manufacturing Engineering',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Physical and Theoretical Chemistry',
'Electrical and Electronic Engineering',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Computational Mathematics',
'Applied Mathematics',
'Computer Science Applications',
'Geometry and Topology',
```

```
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Algebra and Number Theory',
'Computational Mathematics',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Algebra and Number Theory',
'Computational Mathematics',
'Energy Engineering and Power Technology',
'Electrical and Electronic Engineering',
'Chemical Engineering (miscellaneous)',
'Radiology, Nuclear Medicine and Imaging',
'Acoustics and Ultrasonics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
```

```
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Theoretical Computer Science',
'Human-Computer Interaction',
'Industrial and Manufacturing Engineering',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Renewable Energy, Sustainability and the Environment',
'Nuclear Energy and Engineering',
'Fuel Technology',
'Energy Engineering and Power Technology',
'Social Sciences (miscellaneous)',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Numerical Analysis',
'Modeling and Simulation',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Algebra and Number Theory',
'Computational Mathematics',
'Energy Engineering and Power Technology',
'Electrical and Electronic Engineering',
'Energy Engineering and Power Technology',
'Electrical and Electronic Engineering',
'Computer Science Applications',
'Geometry and Topology',
'Control and Optimization',
'Computational Theory and Mathematics',
'Computational Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Information Systems',
'Hardware and Architecture',
'Electronic, Optical and Magnetic Materials',
'Instrumentation',
'Condensed Matter Physics',
'Surfaces, Coatings and Films',
```

```
'Metals and Alloys',
'Electrical and Electronic Engineering',
'Materials Chemistry',
'Software',
'Theoretical Computer Science',
'Human-Computer Interaction',
'Software',
'Theoretical Computer Science',
'Human-Computer Interaction',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Environmental Engineering',
'Environmental Chemistry',
'Waste Management and Disposal',
'Pollution',
'Environmental Engineering',
'Environmental Chemistry',
'Waste Management and Disposal',
'Pollution',
'Statistics and Probability',
'Condensed Matter Physics',
'Control and Systems Engineering',
'Signal Processing',
'Civil and Structural Engineering',
'Aerospace Engineering',
'Mechanical Engineering',
```

```
'Computer Science Applications',
'Analytical Chemistry',
'Atomic and Molecular Physics, and Optics',
'Instrumentation',
'Spectroscopy',
'Biotechnology',
'Biophysics',
'Biomedical Engineering',
'Electrochemistry',
'Behavioral Neuroscience',
'Statistics and Probability',
'Condensed Matter Physics',
'Pharmacology',
'Drug Discovery',
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Condensed Matter Physics',
'Instrumentation',
'Electrical and Electronic Engineering',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Safety, Risk, Reliability and Quality',
'Safety Research',
'Public Health, Environmental and Occupational Health',
'Renewable Energy, Sustainability and the Environment',
'Signal Processing',
'Health Informatics',
'Statistics and Probability',
'Condensed Matter Physics',
'Soil Science',
'Analytical Chemistry',
'Food Science',
'Medicine (all)',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Food Science',
'Applied Microbiology and Biotechnology',
```

```
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science (all)',
'Chemistry (all)',
'Materials Science (all)',
'Mechanics of Materials',
'Physics and Astronomy (all)',
'Computational Mathematics',
'Energy Engineering and Power Technology',
'Electrical and Electronic Engineering',
'Acoustics and Ultrasonics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Safety, Risk, Reliability and Quality',
'Industrial and Manufacturing Engineering',
'Materials Science (all)',
'Condensed Matter Physics',
'Engineering (all)',
'Fluid Flow and Transfer Processes',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
```

```
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Renewable Energy, Sustainability and the Environment',
'Electronic, Optical and Magnetic Materials',
'Instrumentation',
'Condensed Matter Physics',
'Surfaces, Coatings and Films',
'Metals and Alloys',
'Materials Chemistry',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Renewable Energy, Sustainability and the Environment',
'Renewable Energy, Sustainability and the Environment',
'Renewable Energy, Sustainability and the Environment',
'Statistics and Probability',
'Condensed Matter Physics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Statistics and Probability',
'Condensed Matter Physics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Renewable Energy, Sustainability and the Environment',
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Electrical and Electronic Engineering',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Signal Processing',
'Health Informatics',
'Theoretical Computer Science',
```

```
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Theoretical Computer Science',
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Theoretical Computer Science',
'Signal Processing',
'Information Systems',
'Computer Science Applications',
'Computational Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computer Science Applications',
'Computer Graphics and Computer-Aided Design',
'Industrial and Manufacturing Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Theoretical Computer Science',
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Computer Science (all)',
'Modeling and Simulation',
```

```
'Management Science and Operations Research',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Theoretical Computer Science',
'Signal Processing',
'Information Systems',
'Computer Science Applications',
'Theoretical Computer Science',
'Signal Processing',
'Information Systems',
'Computer Science Applications',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Theoretical Computer Science',
'Signal Processing',
'Information Systems',
'Computer Science Applications',
'Software',
```

```
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Hardware and Architecture',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Information Systems',
'Computer Networks and Communications',
'Library and Information Sciences',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Nuclear Energy and Engineering',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
```

```
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Arts and Humanities (miscellaneous)',
'Human-Computer Interaction',
'Psychology (all)',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Renewable Energy, Sustainability and the Environment',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Biophysics',
'Biomedical Engineering',
'Radiology, Nuclear Medicine and Imaging',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Control and Systems Engineering',
'Software',
'Mathematics (all)',
```

```
'Computer Science Applications',
'Industrial and Manufacturing Engineering',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Theoretical Computer Science',
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Software',
'Hardware and Architecture',
'Computer Networks and Communications',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Statistics and Probability',
'Condensed Matter Physics',
'Theoretical Computer Science',
'Discrete Mathematics and Combinatorics',
```

```
'Computational Theory and Mathematics',
'Theoretical Computer Science',
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Instrumentation',
'Electrical and Electronic Engineering',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Computational Mathematics',
'Applied Mathematics',
'Theoretical Computer Science',
'Computer Networks and Communications',
'Computational Theory and Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Control and Systems Engineering',
'Analysis',
'Computer Science Applications',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Hardware and Architecture',
'Electrical and Electronic Engineering',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Engineering (all)',
'Computer Science Applications',
```

```
'Artificial Intelligence',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Chemical Engineering (all)',
'Nuclear Energy and Engineering',
'Aerospace Engineering',
'Mechanical Engineering',
'Fluid Flow and Transfer Processes',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Computer Science (miscellaneous)',
'Computational Mathematics',
'Signal Processing',
'Health Informatics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Signal Processing',
'Health Informatics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Bioengineering',
'Mechanics of Materials',
'Mechanical Engineering',
'Computer Science Applications',
'Energy Engineering and Power Technology',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
```

```
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Mechanical Engineering',
'Electrical and Electronic Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Control and Systems Engineering',
'Signal Processing',
'Civil and Structural Engineering',
'Aerospace Engineering',
'Mechanical Engineering',
'Computer Science Applications',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Control and Systems Engineering',
'Signal Processing',
'Civil and Structural Engineering',
'Aerospace Engineering',
'Mechanical Engineering',
'Computer Science Applications',
'Human Factors and Ergonomics',
'Physical Therapy, Sports Therapy and Rehabilitation',
'Safety, Risk, Reliability and Quality',
'Engineering (miscellaneous)',
'Control and Systems Engineering',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Electrical and Electronic Engineering',
'Engineering (all)',
'Computer Science Applications',
'Artificial Intelligence',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Hardware and Architecture',
'Computer Networks and Communications',
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
```

```
'Computational Mathematics',
'Applied Mathematics',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Modeling and Simulation',
'Management Science and Operations Research',
'Information Systems and Management',
'Control and Systems Engineering',
'Software',
'Mathematics (all)',
'Computer Science Applications',
'Industrial and Manufacturing Engineering',
'Environmental Engineering',
'Environmental Chemistry',
'Waste Management and Disposal',
'Pollution',
'Computational Mathematics',
'Applied Mathematics',
'Computer Science (all)',
'Modeling and Simulation',
'Management Science and Operations Research',
'Software',
'Theoretical Computer Science',
'Hardware and Architecture',
'Computer Networks and Communications',
'Artificial Intelligence',
'Algebra and Number Theory',
'Computational Mathematics',
'Applied Mathematics',
'Computational Mathematics',
'Applied Mathematics',
'Electronic, Optical and Magnetic Materials',
'Atomic and Molecular Physics, and Optics',
'Electrical and Electronic Engineering',
'Software',
'Control and Systems Engineering',
'Theoretical Computer Science',
'Computer Science Applications',
'Information Systems and Management',
'Artificial Intelligence',
'Software',
'Signal Processing',
'Computer Vision and Pattern Recognition',
'Artificial Intelligence',
'Software',
'Signal Processing',
```

```
 'Computer Vision and Pattern Recognition',
 'Artificial Intelligence',
 'Software',
 'Control and Systems Engineering',
 'Theoretical Computer Science',
 'Computer Science Applications',
 'Information Systems and Management',
 'Artificial Intelligence',
 'Control and Systems Engineering',
 'Software',
 'Signal Processing',
 'Computer Vision and Pattern Recognition',
 'Electrical and Electronic Engineering',
 'Engineering (all)',
 'Computer Science Applications',
 'Artificial Intelligence',
 'Nuclear Energy and Engineering',
 'Control and Systems Engineering',
 'Signal Processing',
 'Civil and Structural Engineering',
 'Aerospace Engineering',
 'Mechanical Engineering',
 'Computer Science Applications',
 ...]
```

```
In [41]: from collections import Counter
```

```
In [42]: count = Counter(flat_li)
```

```
In [43]: count
```

```
Out[43]: Counter({'Acoustics and Ultrasonics': 4,
                  'Aerospace Engineering': 18,
                  'Agricultural and Biological Sciences (all)': 1,
                  'Agronomy and Crop Science': 2,
                  'Algebra and Number Theory': 6,
                  'Analysis': 4,
                  'Analytical Chemistry': 6,
                  'Applied Mathematics': 65,
                  'Applied Microbiology and Biotechnology': 1,
                  'Artificial Intelligence': 205,
                  'Arts and Humanities (miscellaneous)': 5,
                  'Atomic and Molecular Physics, and Optics': 13,
                  'Automotive Engineering': 14,
                  'Behavioral Neuroscience': 1,
                  'Biochemistry, Genetics and Molecular Biology (all)': 1,
                  'Bioengineering': 4,
                  'Biomaterials': 1,
                  'Biomedical Engineering': 12,
```

```
'Biophysics': 2,
'Biotechnology': 1,
'Building and Construction': 7,
'Business and International Management': 7,
'Business, Management and Accounting (all)': 1,
'Chemical Engineering (all)': 7,
'Chemical Engineering (miscellaneous)': 2,
'Chemistry (all)': 6,
'Civil and Structural Engineering': 27,
'Cognitive Neuroscience': 13,
'Colloid and Surface Chemistry': 1,
'Communication': 3,
'Computational Mathematics': 65,
'Computational Mechanics': 1,
'Computational Theory and Mathematics': 24,
'Computer Graphics and Computer-Aided Design': 2,
'Computer Networks and Communications': 25,
'Computer Science (all)': 429,
'Computer Science (miscellaneous)': 18,
'Computer Science Applications': 175,
'Computer Vision and Pattern Recognition': 72,
'Condensed Matter Physics': 33,
'Control and Optimization': 30,
'Control and Systems Engineering': 403,
'Decision Sciences (all)': 63,
'Decision Sciences (miscellaneous)': 14,
'Development': 1,
'Developmental and Educational Psychology': 1,
'Discrete Mathematics and Combinatorics': 2,
'Drug Discovery': 1,
'Earth and Planetary Sciences (all)': 2,
'Earth and Planetary Sciences (miscellaneous)': 1,
'Earth-Surface Processes': 1,
'Ecology': 1,
'Ecology, Evolution, Behavior and Systematics': 1,
'Economics and Econometrics': 1,
'Economics, Econometrics and Finance (miscellaneous)': 14,
'Education': 6,
'Electrical and Electronic Engineering': 60,
'Electrochemistry': 1,
'Electronic, Optical and Magnetic Materials': 14,
'Energy Engineering and Power Technology': 22,
'Engineering (all)': 79,
'Engineering (miscellaneous)': 4,
'Environmental Chemistry': 5,
'Environmental Engineering': 5,
'Environmental Science (all)': 1,
'Experimental and Cognitive Psychology': 2,
```

```
'Fluid Flow and Transfer Processes': 4,
'Food Science': 8,
'Fuel Technology': 2,
'Geometry and Topology': 16,
'Geotechnical Engineering and Engineering Geology': 1,
'Hardware and Architecture': 49,
'Health Informatics': 16,
'Health, Toxicology and Mutagenesis': 1,
'Horticulture': 1,
'Human Factors and Ergonomics': 10,
'Human-Computer Interaction': 13,
'Immunology': 1,
'Immunology and Microbiology (all)': 1,
'Industrial and Manufacturing Engineering': 51,
'Infectious Diseases': 1,
'Information Systems': 41,
'Information Systems and Management': 65,
'Instrumentation': 8,
'Language and Linguistics': 2,
'Law': 1,
'Library and Information Sciences': 26,
'Linguistics and Language': 2,
'Management Information Systems': 1,
'Management Science and Operations Research': 44,
'Management of Technology and Innovation': 6,
'Marketing': 1,
'Materials Chemistry': 4,
'Materials Science (all)': 4,
'Mathematics (all)': 8,
'Mechanical Engineering': 44,
'Mechanics of Materials': 14,
'Media Technology': 1,
'Medicine (all)': 1,
'Metals and Alloys': 3,
'Microbiology': 1,
'Microbiology (medical)': 1,
'Modeling and Simulation': 43,
'Neurology': 4,
'Neuroscience (all)': 2,
'Neuroscience (miscellaneous)': 1,
'Nuclear Energy and Engineering': 8,
'Numerical Analysis': 5,
'Ocean Engineering': 1,
'Pharmacology': 1,
'Physical Therapy, Sports Therapy and Rehabilitation': 6,
'Physical and Theoretical Chemistry': 4,
'Physics and Astronomy (all)': 4,
'Pollution': 6,
```

```
                  'Psychology (all)': 6,
                  'Public Health, Environmental and Occupational Health': 1,
                  'Radiological and Ultrasound Technology': 2,
                  'Radiology, Nuclear Medicine and Imaging': 5,
                  'Renewable Energy, Sustainability and the Environment': 21,
                  'Safety Research': 1,
                  'Safety, Risk, Reliability and Quality': 9,
                  'Signal Processing': 128,
                  'Social Sciences (miscellaneous)': 15,
                  'Sociology and Political Science': 1,
                  'Software': 155,
                  'Soil Science': 6,
                  'Spectroscopy': 2,
                  'Statistics and Probability': 29,
                  'Strategy and Management': 7,
                  'Surfaces and Interfaces': 1,
                  'Surfaces, Coatings and Films': 6,
                  'Theoretical Computer Science': 51,
                  'Tourism, Leisure and Hospitality Management': 1,
                  'Toxicology': 1,
                  'Urban Studies': 1,
                  'Waste Management and Disposal': 5})

In [44]: sorted_count = sorted(count.items(), key=operator.itemgetter(1))

In [45]: sorted_count

Out[45]: [('Biotechnology', 1),
          ('Electrochemistry', 1),
          ('Behavioral Neuroscience', 1),
          ('Pharmacology', 1),
          ('Drug Discovery', 1),
          ('Safety Research', 1),
          ('Public Health, Environmental and Occupational Health', 1),
          ('Medicine (all)', 1),
          ('Applied Microbiology and Biotechnology', 1),
          ('Marketing', 1),
          ('Development', 1),
          ('Sociology and Political Science', 1),
          ('Urban Studies', 1),
          ('Tourism, Leisure and Hospitality Management', 1),
          ('Toxicology', 1),
          ('Health, Toxicology and Mutagenesis', 1),
          ('Computational Mechanics', 1),
          ('Biomaterials', 1),
          ('Colloid and Surface Chemistry', 1),
          ('Developmental and Educational Psychology', 1),
          ('Surfaces and Interfaces', 1),
```

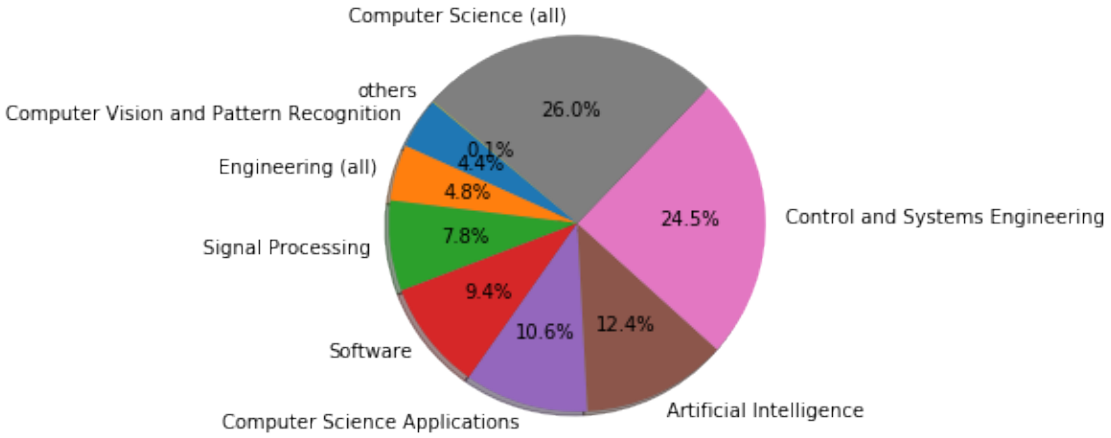
```
('Biochemistry, Genetics and Molecular Biology (all)', 1),
('Immunology and Microbiology (all)', 1),
('Agricultural and Biological Sciences (all)', 1),
('Earth and Planetary Sciences (miscellaneous)', 1),
('Geotechnical Engineering and Engineering Geology', 1),
('Ocean Engineering', 1),
('Microbiology', 1),
('Immunology', 1),
('Microbiology (medical)', 1),
('Infectious Diseases', 1),
('Media Technology', 1),
('Business, Management and Accounting (all)', 1),
('Economics and Econometrics', 1),
('Ecology, Evolution, Behavior and Systematics', 1),
('Ecology', 1),
('Environmental Science (all)', 1),
('Horticulture', 1),
('Neuroscience (miscellaneous)', 1),
('Management Information Systems', 1),
('Earth-Surface Processes', 1),
('Law', 1),
('Discrete Mathematics and Combinatorics', 2),
('Neuroscience (all)', 2),
('Chemical Engineering (miscellaneous)', 2),
('Fuel Technology', 2),
('Spectroscopy', 2),
('Biophysics', 2),
('Computer Graphics and Computer-Aided Design', 2),
('Experimental and Cognitive Psychology', 2),
('Earth and Planetary Sciences (all)', 2),
('Language and Linguistics', 2),
('Linguistics and Language', 2),
('Agronomy and Crop Science', 2),
('Radiological and Ultrasound Technology', 2),
('Metals and Alloys', 3),
('Communication', 3),
('Analysis', 4),
('Physical and Theoretical Chemistry', 4),
('Acoustics and Ultrasonics', 4),
('Materials Chemistry', 4),
('Materials Science (all)', 4),
('Physics and Astronomy (all)', 4),
('Fluid Flow and Transfer Processes', 4),
('Bioengineering', 4),
('Engineering (miscellaneous)', 4),
('Neurology', 4),
('Numerical Analysis', 5),
('Radiology, Nuclear Medicine and Imaging', 5),
```

```
('Environmental Engineering', 5),
('Environmental Chemistry', 5),
('Waste Management and Disposal', 5),
('Arts and Humanities (miscellaneous)', 5),
('Psychology (all)', 6),
('Analytical Chemistry', 6),
('Soil Science', 6),
('Algebra and Number Theory', 6),
('Surfaces, Coatings and Films', 6),
('Pollution', 6),
('Chemistry (all)', 6),
('Physical Therapy, Sports Therapy and Rehabilitation', 6),
('Education', 6),
('Management of Technology and Innovation', 6),
('Chemical Engineering (all)', 7),
('Business and International Management', 7),
('Building and Construction', 7),
('Strategy and Management', 7),
('Food Science', 8),
('Nuclear Energy and Engineering', 8),
('Instrumentation', 8),
('Mathematics (all)', 8),
('Safety, Risk, Reliability and Quality', 9),
('Human Factors and Ergonomics', 10),
('Biomedical Engineering', 12),
('Atomic and Molecular Physics, and Optics', 13),
('Human-Computer Interaction', 13),
('Cognitive Neuroscience', 13),
('Electronic, Optical and Magnetic Materials', 14),
('Mechanics of Materials', 14),
('Decision Sciences (miscellaneous)', 14),
('Economics, Econometrics and Finance (miscellaneous)', 14),
('Automotive Engineering', 14),
('Social Sciences (miscellaneous)', 15),
('Geometry and Topology', 16),
('Health Informatics', 16),
('Aerospace Engineering', 18),
('Computer Science (miscellaneous)', 18),
('Renewable Energy, Sustainability and the Environment', 21),
('Energy Engineering and Power Technology', 22),
('Computational Theory and Mathematics', 24),
('Computer Networks and Communications', 25),
('Library and Information Sciences', 26),
('Civil and Structural Engineering', 27),
('Statistics and Probability', 29),
('Control and Optimization', 30),
('Condensed Matter Physics', 33),
('Information Systems', 41),
```

```
 ('Modeling and Simulation', 43),
 ('Management Science and Operations Research', 44),
 ('Mechanical Engineering', 44),
 ('Hardware and Architecture', 49),
 ('Industrial and Manufacturing Engineering', 51),
 ('Theoretical Computer Science', 51),
 ('Electrical and Electronic Engineering', 60),
 ('Decision Sciences (all)', 63),
 ('Applied Mathematics', 65),
 ('Computational Mathematics', 65),
 ('Information Systems and Management', 65),
 ('Computer Vision and Pattern Recognition', 72),
 ('Engineering (all)', 79),
 ('Signal Processing', 128),
 ('Software', 155),
 ('Computer Science Applications', 175),
 ('Artificial Intelligence', 205),
 ('Control and Systems Engineering', 403),
 ('Computer Science (all)', 429)]
```

In [52]:
```
temp = sorted_count[-8:]

labels = []
sizes = []

for item in temp:
    labels.append(item[0])
    sizes.append(item[1])

labels.append("others")
sizes.append(1)
```

In [54]:
```
# Data to plot
# colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue']
# explode = (0.1, 0, 0, 0)  # explode 1st slice

# Plot
plt.pie(sizes, labels=labels,
        autopct='%1.1f%%', shadow=True, startangle=140)

plt.axis('equal')
plt.show()
```


Computer Science (all)
26.0%
others
Computer Vision and Pattern Recognition
0.1%
Engineering (all)
4.4%
4.8%
Signal Processing
7.8%
9.4%
Software
10.6%
Computer Science Applications
12.4%
Artificial Intelligence
24.5%
Control and Systems Engineering