In [98]:
```
import pandas as pd
from pyzotero import zotero
from IPython.display import display
from sklearn.cluster import AffinityPropagation
import distance
import numpy as np
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000)
```

Zotero add-on

We can extract all tags in a library using the add-on **zotero-tag** by right-clicking on the collection and exporting all tags as csv. It comes out with 4 columns - tag, count, items and ids. The items and ids columns contain lists separated by a comma, which bugs the python parser.
I opened it with LibreOffoce Calc and save it with semi-columns as a separtor.

In [99]:
```
df = pd.read_csv('tags_fixed.csv',sep=';')
df.head()
```

Out[99]:

	TAG	COUNT	ITEMS	IDS
0	/unread	551	A BIM-integrated Fuzzy Multi-criteria Decision...	474,306,163,54,461,580,103,395,493,165,525,127...
1	Building Information Modeling (BIM)	72	Semantic information alignment of BIMs to comp...	306,54,493,128,518,472,15,45,458,482,440,155,5...
2	Automated Compliance Checking (ACC)	30	Semantic information alignment of BIMs to comp...	306,54,15,458,590,448,102,447,58,577,425,600,3...
3	ontology	42	Semantic information alignment of BIMs to comp...	306,493,165,792,128,472,45,482,9,140,288,46,29...
4	Information extraction	2	Semantic information alignment of BIMs to comp...	306159

In [100…
```
tags_addon = df['TAG']
print(tags_addon.size)
```

866

Zotero API

Here is an API which gives more detailed information on the group/personal libraries in JSON format. The library pyzotero helps in the usage. Let's compare the results.

The returned list contains duplicates. Strangely two we get two tags not found from the addon(mby set transforamtion)

In [101…
```
# API key is personal, The id is of the grou[]
zot = zotero.Zotero(3007408, 'group', 'Yv2xY0CH9vjjf40sQiVcCQt3')
tags_api = pd.Series(zot.everything(zot.tags()))
```

In [102…
```
len(tags_api)
```

Out[102]: 921

In [103…
```
tags_api.duplicated().sum()
```

Out[103]: 55

In [104…
```
tags_addon.duplicated().sum()
```

Out[104]: 0

In [105…
```
dif = set(tags_api.tolist()) - set(tags_addon.tolist())
print(dif)
```

```
{'Object-Property, Method, Relation', 'Rule language, engine and checking'}
```

Here is an example of a single item from the library and it's metadata

In [106…
```
items = zot.items()
print(items[0])
```

{'key': 'IHC2TIKN', 'version': 1987, 'library': {'type': 'group', 'id': 300
7408, 'name': 'Semantic BIM', 'links': {'alternate': {'href': 'https://www.
zotero.org/groups/semantic_bim', 'type': 'text/html'}}}, 'links': {'self':
{'href': 'https://api.zotero.org/groups/3007408/items/IHC2TIKN', 'type': 'a
pplication/json'}, 'alternate': {'href': 'https://www.zotero.org/groups/sem
antic_bim/items/IHC2TIKN', 'type': 'text/html'}}, 'meta': {'createdByUser':
{'id': 10307964, 'username': 'aleksandrositsyn', 'name': 'Aleksandr Ositsy
n', 'links': {'alternate': {'href': 'https://www.zotero.org/aleksandrositsy
n', 'type': 'text/html'}}}, 'creatorSummary': 'Corry et al.', 'parsedDate':
'2014-10', 'numChildren': 0}, 'data': {'key': 'IHC2TIKN', 'version': 1987,
'itemType': 'journalArticle', 'title': 'Using semantic web technologies to
access soft AEC data', 'creators': [{'creatorType': 'author', 'firstName':
'Edward', 'lastName': 'Corry'}, {'creatorType': 'author', 'firstName': 'Jam
es', 'lastName': 'O'Donnell'}, {'creatorType': 'author', 'firstName': 'Edwa
rd', 'lastName': 'Curry'}, {'creatorType': 'author', 'firstName': 'Daniel',
'lastName': 'Coakley'}, {'creatorType': 'author', 'firstName': 'Pieter', 'l
astName': 'Pauwels'}, {'creatorType': 'author', 'firstName': 'Marcus', 'las
tName': 'Keane'}], 'abstractNote': 'Building related data tends to be gener
ated, used and retained in a domain-specific manner. The lack of interopera
bility between data domains in the architecture, engineering and constructi
on (AEC) industry inhibits the cross-domain use of data at an enterprise le
vel. Semantic web technologies provide a possible solution to some of the n
oted interoperability issues. Traditional methods of information capture fa
il to take into account the wealth of soft information available throughout
a building. Several sources of information are not included in performance
assessment frameworks, including social media, occupant communication, mobi
le communication devices, occupancy patterns, human resource allocations an
d financial information. The paper suggests that improved data interoperabi
lity can aid the integration of untapped silos of information into existing
structured performance measurement frameworks, leading to greater awareness
of stakeholder concerns and building performance. An initial study of how b
uilding-related data can be published following semantic web principles and
integrated with other 'soft-data' sources in a cross-domain manner is prese
nted. The paper goes on to illustrate how data sources from outside the bui
lding operation domain can be used to supplement existing sources. Future w
ork will include the creation of a semantic web based performance framework
platform for building performance optimisation.', 'publicationTitle': 'Adva
nced Engineering Informatics', 'volume': '28', 'issue': '4', 'pages': '370–
380', 'date': '2014-10', 'series': '', 'seriesTitle': '', 'seriesText': '',
'journalAbbreviation': '', 'language': 'en', 'DOI': '10.1016/j.aei.2014.05.
002', 'ISSN': '1474-0346', 'shortTitle': '', 'url': 'https://www.sciencedir
ect.com/science/article/pii/S1474034614000366', 'accessDate': '2021-05-14',
'archive': '', 'archiveLocation': '', 'libraryCatalog': '', 'callNumber':
'', 'rights': '', 'extra': 'Citation Key: corryUsingSemanticWeb2014', 'tag
s': [{'tag': '/unread'}, {'tag': 'Building performance'}, {'tag': 'Linked d
ata'}, {'tag': 'Performance metrics'}, {'tag': 'RDF'}, {'tag': 'Social medi
a'}, {'tag': 'Twitter'}], 'collections': [], 'relations': {}, 'dateAdded':
'2022-10-12T14:04:27Z', 'dateModified': '2022-10-12T14:04:27Z'}}

Normalization

Even though Vlado suggested to use the algo's present in openrefine, a found a string clustering algo in sklearn library based on levenstein distance. The algorith gives a warning of failed convergence thus results may not be optimal.

In [107…

```
# Using the tags from the addon plugin
words = np.asarray(tags_addon)
lev_similarity = -1*np.array([[distance.levenshtein(w1,w2) for w1 in words]
```

```
affprop = AffinityPropagation(affinity="precomputed", damping=0.5)
affprop.fit(lev_similarity)
for cluster_id in np.unique(affprop.labels_):
    exemplar = words[affprop.cluster_centers_indices_[cluster_id]]
    cluster = np.unique(words[np.nonzero(affprop.labels_==cluster_id)])
    cluster_str = ", ".join(cluster)
    print(" - *%s:* %s" % (exemplar, cluster_str))
```

- *ontology:* Airport ontology, B-Prolog, Basic Formal Ontology, Bozen-Bolzano, CQIEOntology, CSCOntology, EUnet4DBP, Energy, Geology, IndoorGML, Meteorology, Prolog, Topology, Vocabulary, energy, ontology, ontology mapping, point cloud, pythonOCC, review methodology, taxonomy, terminology
- *Visual Compliance Checking Language:* Visual Code Checking Language, Visual Compliance Checking Language
- *spatial reasoning:* Declarative Reasoning, as-designed/as-built, defeasible reasoning, logic-based reasoning, spatial computation, spatial indexing, spatial operators, spatial reasoning, topological reasoning
- *International Foundation Classes (IFC):* International Foundation Classes (IFC)
- *Model view definition (MVD):* Model View Definition (MVD), Model view definition (MVD), Model view definitions (MVD)
- *Hazard recognition and communication:* Hazard recognition and communication
- *Occupational construction safety and health:* Occupational construction safety and health
- *Schema:* AC Demat, Benchmark, BrickSchema, Florida, Geneva, Korea, Occupancy, Roadmap, Rotterdam, SEUMTER, SPIN Jena, Schema, Snøhetta, Social media, assembly, benchmark, ifcJSON
- *Smart Buildings:* Smart Building, Smart Buildings, SmartCodes Builder, near-zero energy buildings, smart buildings
- *Common Data Environenment (CDE):* Common Data Environenment (CDE), Common Data Environment (CDE)
- *Sustainable Development Goals (SDG):* Sustainable Development Goals (SDG)
- *DAta Linked Through Occurrences Network:* DAta Linked Through Occurrences Network
- *Formalisation of conformance requirements:* Formalisation of conformance requirements
- *Ontological approach for conformance checking:* Ontological approach for conformance checking
- *Semantic annotation and organisation of building codes:* Semantic annotation and organisation of building codes
- *information extraction:* Automated information extraction, Information extraction, Information retrieval, information container, information extraction
- *Compliance Audit Procedures:* Compliance Audit Procedures, Compliant design procedures, Visual Audit Procedures
- *Legal Knowledge Model:* Legal Knowledge Model, Regulatory knowledge model
- *SPARQL:* BimSPARQL, GeoSPARQL, IfcOWL, SAREF, SAREF4BLDG, SHACL, SPARQL, SPARQL Construct, SPARQL complexity, SPARQL-generate, SWRL, mvdXML
- *Information Container for Document Delivery (ICDD):* Information Container for Document Delivery (ICDD)
- *ISO 21597:* BS/EN 16310, ISO 12006, ISO 12006-2, ISO 16739-1, ISO 19650, ISO 21597, ISO 21597-1, ISO 21838-2, ISO 22263, ISO 23387, ISO16739, NIST Tariff
- *geometry:* Bimfactory, Geodata, Geometry, Geometry rule, Industry 4.0, Offset geometry, Solid geometry, comfort, country, geodata, geometry
- *Level of Detail (LOD):* Level of Detail (LOD), Level of Geometry (LOG), Level of Information Need (LOIN), Level of Interest (LOI), Levels of Detail
- *conflict-driven constraint learning:* conflict-driven constraint learning
- *easoner:* /best-paper, BauVorIV, Bauantrag, Bossam reasoner, Demand Response, Engines, Estonia, Haystack, Hyperledger, LandInfra, LegalDocML, N3Logic, Navisworks, Pareto set, RealEstateCore, Reasoning, Singapore, Web-viewer, designbuilder, easoner, ecosystem, environmental, lessons learned, neural network, temporal
- *Performance-based Building Code:* New Zealand Building Code, Performance-based Building Code

- *Built Environment:* Built Environment, Open Government, Runtime Environment
- *Concepts:* Agents, Akoma Ntoso, Checking, CityGML 3.0, Concepts, Conceptual Graph, Concrete, Context, Contexts, Processes
- *Technologies:* Methodologies, Semantic Web Technologies, Technologies, Web technologies, decentralized, search engine
- *building administration permission service:* building administration permission service
- *energy performance certificates:* energy performance certificates, energy performance certification
- *Building management systems:* BIM management system, Building management systems, building energy management systems
- *Fault detection and diagnostics:* Fault detection and diagnostics
- *renovation:* Automation, Documentation, Function, Geolocation, Information, Integration, OpenCDE-API, Pre-evaluation, Process optimization, Query Execution, Regulations, Verification, event-driven, firefighting, information, innovation, renovation, renovation process, rule extraction, segmentation
- *‘passing’ and ‘failing’ concepts:* ‘passing’ and ‘failing’ concepts
- *business case for use of virtual construction:* business case for use of virtual construction
- *future of virtual construction and regulation checking:* future of virtual construction and regulation checking
- *Solibri Model Checker:* EDM Model Checker, Solibri Model Checker
- *BIM Service:* BIM Service, BIM querying, BIM server, BIM4REN project, Oriented Service, Shared Service, microservices
- *Data integration:* 3D land administration, Continuous Integration, Data Dictionaries, Data integration, Data transformation, Dutch Rgd BIM Norm, Intelligent evaluation, Layout generation, Metadata annotation, WoT Thing Description, data interchange, ifcOWL instantiation, integration
- *Data sharing:* Annual Saving, Blockchain, Data quality, Data sharing, Deep Learning, Deep learning, Dubai GIS Center, Laser scanning, SWRLJessBridge, Safety planning, additive manufacturing, back-jumping, lazy grounding, manufacturing, occupant behavior, path search
- *Rules:* /unread, CORENET, Dubai, EXPRESS rules, FME-desktop, Finland, Fuzzy, Jena Rules, LegalRuleML, Malaysia, Milan, PROV-O, Quality, RCC-8, RDF-star, Ratio rule, Rules, Seumter, Solid, Tools, Turkey, Utilities, deltas, fuzzy sets, real estate
- *Netherlands:* Netherlands, New Zealand, Performance, Requirements, Switzerland, power plants, provenance
- *BIM Rule Language (BIMRL):* BIM Query Language (BIMQL), BIM Rule Language (BIMRL)
- *BIMRL Simplified Schema (BIMRLSS):* BIMRL Simplified Schema (BIMRLSS)
- *Information management:* Asset Information Requirements, Construction management, Information management, Performance management, Semantic information alignment, construction management
- *Digital construction:* Digital Construction, Digital construction, building and construction, design by constraints
- *Construction engineering:* Architectural engineering, Civil engineering, Communication in engineering, Computing in civil engineering, Construction engineering, Construction technical plan
- *building automation:* Building Automation, Building Automation Systems, Building Digital Twin, Building Regulations, Building Topology Ontology, Formalisation automatique, Green building evaluation, bill of quantities, building acoustics, building automation, building benchmarking, building circulation, engineering optimisation
- *Rule language:* BERA language, KBim Visual Language, Query language, Rule engine, Rule language, Structured Natural Language, executable language
- *Construction:* BIM object recognition, Computer applications, Constraint, Constructability, Construction, Construction 4.0, Cost estimate, IfcConstraint, change detection, construction

- *Linked data:* Linked Data, Linked Data Platform (LDP), Linked data, Linked open data, Numeric data, Time-series data, big data, linked data, metadata, planned/actual
- *Semantic interoperability:* Semantic Interoperability, Semantic interoperability
- *Wireless sensor networks:* Water distribution networks, Wireless Sensor Networks, Wireless sensor networks
- *energy management:* Building Management, Energy management, Knowledge Management, Knowledge management, Onlinezugangsgesetz, Project management, Site management, Temperature measurement, Use Case Management, asset management, energy management, ontology enrichment
- *ontology and inference rule based framework:* ontology and inference rule based framework
- *Computer Science - Computers and Society:* Computer Science - Computers and Society
- *semantic modeling:* Autonomic Computing, Knowledge modeling, Object oriented modeling, Ontology modelling, Semantic knowledge, Semantic modeling, parametric modeling, process modeling, semantic mapping, semantic modeling, semantic technology
- *ambient intelligence:* Artificial Intelligence, ambient intelligence
- *domain ontologies:* Domain Ontology, Fuzzy ontologies, Product Catalogue, domain ontologies, engineering ontologies, enrichissement d'ontologie, spatial ontologies
- *Electricity:* Electricity, Electricity Market, Electronic books, air quality, uncertainty
- *Semantics:* Communities, Performance metrics, Semantic, Semantic BIM, Semantic Web, Semantic web, Semantics, Sensors, Smart City, Smart Grid, Soft Skills, e-maintenance, ePlanCheck, semantics
- *Buildings:* Buildability, Buildings, Built2Spec, Guidelines, Web of Things, building, oil and gas, proceedings
- *Expert System:* Central Server, Energy Systems, Expert System, Expert systems, Export test, Import test, legality system
- *code compliance checking:* Automated Compliance Checking (ACC), Compliance checking, Design compliance checking, code compliance checking, code compliance checks, compliance checking, safety compliance checking
- *Automated design assessment:* Automated design assessment, automated design review
- *Computer-executable form:* Computer-executable form
- *KBimCode:* BimCheck, KBimAssess, KBimCode, KBimCodeDB, KBimLogic, KBimVeri, SimModel, SmartCodes
- *Generative design:* Generative design, Generative interior design, structural design
- *Multiobjective performance optimisation:* Building performance optimisation, Multiobjective performance optimisation
- *interoperability:* Data interoperability, Interoperability, interoperability
- *Data models:* 3D city model, 3D modelling, ASHVIN project, Adaptation models, Data Model, Data models, Facility data model, LOD levels, Load modeling, Management methods, Mathematical model, Maturity Model, Object model, coding model, material properties, safety and health, smart meters
- *Context-free grammar (CFG):* Context-free grammar (CFG)
- *Request for proposal (RFP):* Request for proposal (RFP)
- *Rule interpretation:* Rule interpretation, Rules Classification, rule interpretation
- *AIM:* 3CIM, ACABIM, AEC, AECO, AI, AIM, ASP4BIM, Austria, BIMSeek, BIMbots, BPMN, C/VM2, CityGML, EHR, ELASSTIC, ESRI, EYE, FIBO, FSGIM, GeoBIM, IDM, Jess, KOM, LKM, MBO, MEP, MVD, McBIM, NRM2, NZBC, ONTOP, OPM, Open BIM, Org, POMI2, PSD, QL4BIM, RKM, SITG, SMC, SPIN, TIO, TNO, UCM, UoM, XBIM, eeBIM
- *validation:* BIM collaboration, Design model validation, Digital Twin, Digitalization, HVAC design, IFC Validation, Saudi Arabia, Validation, appl

ication, classification, data validation, fault detection, handbook, valida
tion, visualization
 - *Test-Driven Development:* Behavior-Driven Development, Research and dev
elopment, Test-Driven Development
 - *Long short-term memory (LSTM):* Long short-term memory (LSTM)
 - *Regulatory compliance:* Proof of compliance, Regulation constraint, Reg
ulatory compliance, Semantics of Regulatory Compliance, Urban planning comp
liance
 - *Domain Specific Language:* Domain Specific Language, domain specific la
nguage
 - *Norwegian Statsbygg BIM Manual:* Norwegian Statsbygg BIM Manual
 - *rule checking:* Code Checking Rules, Code checking, Design rule checkin
g, Firewalls (computing), IFC Checker, IFC schema checking, Quality check,
Rule checking, Rule-based checking, Safety checking, Value Stream Mapping,
e-checking, legality checking, model checking, pre-checking, rule checking,
rule complexity, rule reasoning, rule-checking
 - *Machine Learning:* Layout planning, Machine Learning, Production planni
ng, Reinforcement Learning, ifcwebserver.org, machine learning, search spac
e pruning
 - *Natural language processing:* Natural Language Processing, Natural lang
uage processing, Natural language processing (NLP)
 - *Evacuation regulation:* Evacuation regulation, Procedural code generati
on, Semantics of regulations, Utility regulations, construction regulation
s, legislation and regulations
 - *High-rise and complex buildings:* High-rise and complex buildings
 - *Energy analysis:* Bibliometric Analysis, Energy analysis, Energy consum
ption, Energy systems analysis, Indoor spatial analysis, Query expansion, e
nergy saving, logical analysis, sentence analysis, structural analysis
 - *location breakdown structure:* location breakdown structure
 - *Stages:* ByggNett, Entities, Harpaceas, MetaLex, Metadata, Petri nets,
Protege, Siemens, Software, Spatial DBs, Stages, Standards, Stardog, Statsb
ygg, Sweden, Switches, XPlanung, damage, damages, octree, updates
 - *rail:* ArchiCAD, Australia, Brasil, COBie, Drools, France, Germany, Ghe
rkin, GraphDB, GraphQL, IFCowl, Italy, Izmir, Japan, Materials, Neo4J, Nord
ic, Norway, Portugal, Solibri, Sud-Tirol, Variables, graph, have, hybrid, r
ail, review, revit, road
 - *building permit:* Building Permit, Building Permits, Building performan
ce, Building permit, Well-defined Text, building code, building life-cycle,
building materials, building permission, building permit, building permit c
hecking, building permit process, building permits, buildingSmart, explaina
ble CPS
 - *dice:* Bosch, China, Clingo, Diff, IfcDoc, IfcView, India, LicA, Lifecy
cle, Logic, Logic rule, OWL-Time, Twitter, design, dica, dicbm, dicc, dice,
dices, dici, dicl, diclvl, dicob, dicp, dicstg, dicu, dicv, digiKOM, distri
buted, efficiency, lifts
 - *Geographic Information System:* Geographic Information System
 - *Local context analysis (LCA):* Local context analysis (LCA)
 - *Geospatial data:* 3D cadastral maps, Geospatial analysis, Geospatial da
ta, Relational database, geospatial data
 - *DL:* 3D, CG, DALTON, DL, DTDL, DevOps, DiCon, GML, HDF5, ICDD, KBVL, MV
DXML, NDBC, NLP, OWL, PLM, QUDT, RDF, SNL, SQL, SSN, STL, Tx3, VCCL, VPL, b
SDD, ifcOWL
 - *Semantic frame:* Explanation of proof, Semantic Data, Semantic Mediatio
n, Semantic Web Service, Semantic alignment, Semantic analysis, Semantic en
richment, Semantic frame, Semantic framework, Semantic systems, relation ph
rase, semantic data, semantic rules
 - *Energy efficiency:* Energy Efficiency, Energy efficiency, building ener
gy efficiency, energy efficiency
 - *e-Submission system:* Virtual Permitting System, building e-Submission
system, cyber-physical systems, e-Submission, e-Submission system
 - *BAS:* BACS, BAS, BCF, BEMS, BERA, BFO, BMS, BOT, BRMS, BauGB, CAP, CMI

S, EBPD, FOAF, FORNAX, GATE, GIS, Geo, HOAI, IFC, IoT, JESS, JSDAI, KBim, KPI, LBD, LBauO, LPG, OpenADR, RASE, RETE, RIBA, SBVR, SNACC, SOSA, WGS84, X3D, XBau
 - *Answer Set Programming:* Answer Set Programming, Answer Set Programming (ASP), visual programming
 - *Automated construction management systems:* Automated construction management systems
 - *open standards:* BIM standards, Model Based Approach, Open CASCADE, Open Data Standards, open standards
 - *Digital building permit:* Digital building permit, Digital building twin (DBT), Estonian Building Registry
 - *Industry Foundation Classes:* Industry Foundation Classes, Industry Foundation Classes (IFC), Industry foundation classes
 - *Test Indicator Objective:* Test Indicator Objective
 - *transcribe-transform-transfer:* transcribe-transform-transfer
 - *Automatic formalization:* Automatic formalization, Automatic validation, Information formalization, Knowledge formalisation
 - *conversion:* BIM Vision, Factory design, IFC-OWL conversion, Lower Saxony, Normative provisions, OWL class expression, comparison, computer vision, conflict resolution, conversion, underground, ⛔ No DOI found
 - *KBimCode Editor:* KBimAssess-Lite, KBimCode Composer, KBimCode Editor, KBimCode Object Model, KBimSubmission, Limestone calculator
 - *Object-Property$COMMA$ Method$COMMA$ Relation:* Object-Property$COMMA$ Method$COMMA$ Relation
 - *Information Delivery Manual (IDM):* Information Delivery Manual (IDM), Information Delivery Specification (IDS)
 - *Product Data Templates (PDT):* Product Data Templates (PDT)
 - *Collaborative ontology development:* Collaborative ontology development
 - *National Mapping and Cadastral Agencies:* National Mapping and Cadastral Agencies
 - *Environmental Performance Calculation:* Environmental Performance Calculation
 - *Building environmental monitoring:* Building Environment Rule and Analysis, Building environmental monitoring
 - *Knowledge-based system:* Knowledge based systems, Knowledge-based system
 - *Property Sets:* IFC Property templates, Property Sets, Property data, Property ownership, property valuation
 - *Masterplan Digitalisierung:* Masterplan Digitalisierung
 - *Java Standard Data Access Interface:* Java Standard Data Access Interface
 - *use case:* Business rule, PLR-cadastre, Risk assessment, Rule tables, SQL rule, Units of Measure, Use case, fire safety, noun phrase, open access, unit test, update patches, use case, verb phrase
 - *CLML (Crown Legislation Markup Language):* CLML (Crown Legislation Markup Language)
 - *LKIF (Legal Knowledge Interchange Format):* LKIF (Legal Knowledge Interchange Format)
 - *Architecture:* Architecture, Attribute rule, architecture, infrastructure, product cycle
 - *Business rule management system (BRMS):* Business rule management system (BRMS)
 - *Construction site layout planning:* Construction site layout planning
 - *Rule language$COMMA$ engine and checking:* Rule language$COMMA$ engine and checking
 - *Building Information Modeling:* Building Information Modeling, Building Information Modeling (BIM), Building compliance model
 - *Construction safety:* Bidirectional LSTM, Construction industry, Construction methods, Construction project, Construction quality, Construction resource, Construction safety, Construction workflow, Design for safety, construction lifecycle, construction process, well-structured documents

```
 - *decision support:* decision support, decision table, description logic, group decision making, life-cycle design support, version history
 - *Inference rules:* Conference Proceedings, Fire escape route, Inference Engine, Inference rules, Pattern-matching rules, knowledge graphs
 - *Municipality Housing and Zoning:* Municipality Housing and Zoning
```

```
/home/boyan/.local/lib/python3.10/site-packages/sklearn/cluster/_affinity_propagation.py:236: ConvergenceWarning: Affinity propagation did not converge, this model may return degenerate cluster centers and labels.
  warnings.warn(
```

In [108…

```
clusters = [words[center] for center in affprop.cluster_centers_indices_]
clusters
```

Out[108]:
```
['ontology',
 'Visual Compliance Checking Language',
 'spatial reasoning',
 'International Foundation Classes (IFC)',
 'Model view definition (MVD)',
 'Hazard recognition and communication',
 'Occupational construction safety and health',
 'Schema',
 'Smart Buildings',
 'Common Data Environenment (CDE)',
 'Sustainable Development Goals (SDG)',
 'DAta Linked Through Occurrences Network',
 'Formalisation of conformance requirements',
 'Ontological approach for conformance checking',
 'Semantic annotation and organisation of building codes',
 'information extraction',
 'Compliance Audit Procedures',
 'Legal Knowledge Model',
 'SPARQL',
 'Information Container for Document Delivery (ICDD)',
 'ISO 21597',
 'geometry',
 'Level of Detail (LOD)',
 'conflict-driven constraint learning',
 'easoner',
 'Performance-based Building Code',
 'Built Environment',
 'Concepts',
 'Technologies',
 'building administration permission service',
 'energy performance certificates',
 'Building management systems',
 'Fault detection and diagnostics',
 'renovation',
 '‘passing’ and ‘failing’ concepts',
 'business case for use of virtual construction',
 'future of virtual construction and regulation checking',
 'Solibri Model Checker',
 'BIM Service',
 'Data integration',
 'Data sharing',
 'Rules',
 'Netherlands',
 'BIM Rule Language (BIMRL)',
 'BIMRL Simplified Schema (BIMRLSS)',
 'Information management',
 'Digital construction',
 'Construction engineering',
 'building automation',
 'Rule language',
 'Construction',
 'Linked data',
 'Semantic interoperability',
 'Wireless sensor networks',
 'energy management',
 'ontology and inference rule based framework',
 'Computer Science - Computers and Society',
 'semantic modeling',
 'ambient intelligence',
 'domain ontologies',
 'Electricity',
```

```
'Semantics',
'Buildings',
'Expert System',
'code compliance checking',
'Automated design assessment',
'Computer-executable form',
'KBimCode',
'Generative design',
'Multiobjective performance optimisation',
'interoperability',
'Data models',
'Context-free grammar (CFG)',
'Request for proposal (RFP)',
'Rule interpretation',
'AIM',
'validation',
'Test-Driven Development',
'Long short-term memory (LSTM)',
'Regulatory compliance',
'Domain Specific Language',
'Norwegian Statsbygg BIM Manual',
'rule checking',
'Machine Learning',
'Natural language processing',
'Evacuation regulation',
'High-rise and complex buildings',
'Energy analysis',
'location breakdown structure',
'Stages',
'rail',
'building permit',
'dice',
'Geographic Information System',
'Local context analysis (LCA)',
'Geospatial data',
'DL',
'Semantic frame',
'Energy efficiency',
'e-Submission system',
'BAS',
'Answer Set Programming',
'Automated construction management systems',
'open standards',
'Digital building permit',
'Industry Foundation Classes',
'Test Indicator Objective',
'transcribe-transform-transfer',
'Automatic formalization',
'conversion',
'KBimCode Editor',
'Object-Property$COMMA$ Method$COMMA$ Relation',
'Information Delivery Manual (IDM)',
'Product Data Templates (PDT)',
'Collaborative ontology development',
'National Mapping and Cadastral Agencies',
'Environmental Performance Calculation',
'Building environmental monitoring',
'Knowledge-based system',
'Property Sets',
'Masterplan Digitalisierung',
'Java Standard Data Access Interface',
```

```
 'use case',
 'CLML (Crown Legislation Markup Language)',
 'LKIF (Legal Knowledge Interchange Format)',
 'Architecture',
 'Business rule management system (BRMS)',
 'Construction site layout planning',
 'Rule language$COMMA$ engine and checking',
 'Building Information Modeling',
 'Construction safety',
 'decision support',
 'Inference rules',
 'Municipality Housing and Zoning']
```

In [95]:
```
len(clusters)
```

Out[95]: 138