**KEYWORDS**

**Cu Spacy si YAKE**

max\_ngram = 2

deduplication\_threshold = 0.2

keywords\_nr = 10

remove\_pos = ['ADV', 'PRON', 'CCONJ', 'PUNCT', 'PART', 'DET', 'SPACE', 'NUM', 'SYM', 'PROPN']

stop\_words = ['paper', 'present', 'propose']

Pt author\_id = 534:  
('distributed systems', 6.486297518333773e-07)

('opportunistic networks', 2.542880091946773e-06)

('mobile device', 6.326549047884701e-06)

('data', 8.747530629655898e-06)

('network traffic', 1.59438310959238e-05)

('time information', 2.19122143554771e-05)

('cloud computing', 3.602812885925975e-05)

('system architectures', 5.3703830710534795e-05)

('scheduling algorithm', 5.4045301229162336e-05)

('large', 5.636785745618369e-05)

Fara remove\_pos si keywords\_nr = 5:

('distributed systems', 3.1019707051035056e-07)

('opportunistic networks', 1.371932451167339e-06)

('cloud computing', 3.1212460966845947e-06)

('mobile device', 3.4138318996966906e-06)

('network traffic', 8.446037204843576e-06)

max\_ngram = 2

deduplication\_threshold = 0.3

keywords\_nr = 10

remove\_entities = ['PERSON', 'NORP', 'FAC', 'ORG', 'GPE', 'LOC', 'LANGUAGE', 'DATE', 'TIME', 'PERCENT', 'MONEY',\

'QUANTITY', 'CARDINAL', 'ORDINAL', 'PRODUCT']

stop\_words = ['paper', 'present', 'propose', 'datum', 'people', 'result', 'solution', 'case', 'order', 'base', 'ieee', 'privacy', 'policy']

Pt id = 534:

('distributed systems', 3.047774550077539e-07)

('large scale', 8.960567313303865e-07)

('mobile devices', 8.979313987508377e-07)

('Opportunistic networks', 9.832146871223685e-07)

('scale distributed', 1.3407303299565907e-06)

('Cloud Computing', 1.9639352031621148e-06)

('network resources', 7.662840548406415e-06)

('simulation model', 8.899618143067476e-06)

('scheduling algorithms', 9.26431491230422e-06)

('system technologies', 1.0757253324342725e-05)

max\_ngram = 2

deduplication\_threshold = 0.2

keywords\_nr = 5

('distributed systems', 3.047774550077539e-07)

('Opportunistic networks', 9.832146871223685e-07)

('Cloud Computing', 1.9639352031621148e-06)

('mobile device', 4.111580615332783e-06)

('scheduling algorithms', 9.26431491230422e-06)

Pt id = 841:

('microscopy SEM', 4.2898604665759417e-07)

('ray diffraction', 5.928584998620758e-07)

('composite materials', 6.333985159556437e-07)

('drug delivery', 1.8760749965600923e-06)

('thermal analysis', 3.284227725936979e-06)

Pt id = 1284:

('natural language', 7.612217798879714e-07)

('language processing', 1.3494299505050048e-06)

('chat conversations', 2.339661415569574e-06)

('polyphonic model', 7.994775768743788e-06)

('learning techniques', 1.08443263615344e-05)

Cu max\_ngram=3:

('natural language processing', 1.5457826852342103e-08)

('collaborative learning CSCL', 5.933161522158157e-08)

('learning CSCL environments', 2.7068127137640227e-07)

('automated analysis system', 4.6305514615840113e-07)

('textual complexity indices', 5.157311547086552e-07)

Pt id = 562:

('Cloud computing', 5.20287292452525e-07)

('distributed systems', 1.073780008565533e-06)

('scheduling algorithms', 2.9452607330436414e-06)

('Grid environments', 1.5501663729161587e-05)

('nowadays Cloud', 1.7321709507399917e-05)

Pt id = 1284:

('natural language', 7.612217798879714e-07)

('language processing', 1.3494299505050048e-06)

('chat conversations', 2.339661415569574e-06)

('polyphonic model', 7.994775768743788e-06)

('learning techniques', 1.08443263615344e-05)

Pt id = 584:

('language models', 3.371566513405436e-06)

('natural language', 9.739159390461209e-06)

('deep learning', 2.69609715871529e-05)

('model students', 3.753705281806862e-05)

('automated text', 5.234915662209848e-05)

Pt id = 872:

('neural networks', 1.7430880205276488e-06)

('control system', 2.5896852356877334e-06)

('image processing', 2.8179481095091054e-06)

('process control', 5.760868421795922e-06)

('Unmanned Aerial', 9.571797529239613e-06)

Pt id = 1146:

('virtual reality', 6.294100538988312e-07)

('Online virtual', 6.627989851002342e-06)

('learning process', 1.379728542804732e-05)

('game learning', 3.0349014442316856e-05)

('computer games', 3.066744143659512e-05)

**TOPIC MODELLING – LDA**

Pastrand doar adjective si substantive si doar bigrame si trigrame:

remove\_pos = ['ADV', 'PRON', 'PART', 'DET', 'SPACE', 'NUM', 'SYM', 'ADP', 'VERB', 'CCONJ']

stop\_words = ['paper', 'present', 'propose', 'datum', 'people', 'result', 'solution', 'case', 'order', 'base', 'ieee', 'privacy', 'policy', 'new', 'old', 'context']

remove\_entities = ['PERSON', 'NORP', 'FAC', 'ORG', 'GPE', 'LOC', 'LANGUAGE', 'DATE', 'TIME', 'PERCENT', 'MONEY', 'QUANTITY', 'CARDINAL', 'ORDINAL']

lda\_model = LdaMulticore(corpus=corpus, id2word=dictionary, iterations=500, num\_topics=1, workers=4, passes=100)

Author\_id = 562:

0: 0.014\*"cloud computing" + 0.014\*"scheduling algorithm" + 0.010\*"real time" + 0.010\*"resource management" + 0.010\*"large scale" + 0.008\*"satellite image" + 0.006\*"service level" + 0.006\*"task scheduling" + 0.006\*"smart city" + 0.006\*"cloud service"

Author\_id = 534:

0: 0.011\*"large scale" + 0.011\*"mobile device" + 0.008\*"opportunistic network" + 0.005\*"real time" + 0.004\*"cloud computing" + 0.004\*"wide range" + 0.003\*"quality life" + 0.003\*"scale system" + 0.003\*"inconvenience helpful" + 0.003\*"sale account"

Author\_id = 1146:

0: 0.015\*"virtual reality" + 0.008\*"real time" + 0.007\*"virtual environment" + 0.005\*"sensory substitution" + 0.005\*"virtual space" + 0.004\*"sound localization" + 0.004\*"sound source" + 0.004\*"smith chart" + 0.003\*"virtual world" + 0.003\*"augmented reality"

Author\_id = 841:

0: 0.012\*"x ray" + 0.008\*"electron microscopy" + 0.006\*"composite material" + 0.006\*"drug delivery" + 0.005\*"ray diffraction" + 0.004\*"iron oxide" + 0.004\*"sol gel" + 0.003\*"microscopy sem" + 0.003\*"transmission electron" + 0.003\*"thin film"

Author\_id = 872:

0: 0.014\*"neural network" + 0.009\*"real time" + 0.008\*"wireless sensor" + 0.008\*"fractal dimension" + 0.007\*"unmanned aerial" + 0.006\*"sensor network" + 0.006\*"image processing" + 0.006\*"large scale" + 0.005\*"convolutional neural" + 0.005\*"control system"

Author\_id = 1284:

0: 0.011\*"natural language" + 0.008\*"e learning" + 0.006\*"textual complexity" + 0.006\*"chat conversation" + 0.005\*"social network" + 0.004\*"learning process" + 0.003\*"polyphonic model" + 0.003\*"machine learning" + 0.003\*"artificial intelligence" + 0.003\*"web page"

Cu *stop\_words = ['paper', 'present', 'propose', 'datum', 'people', 'result', 'solution', 'case', 'order', 'base', 'ieee', 'privacy', 'policy',*

*'new', 'old', 'context', 'high', 'different', 'new', 'old', 'research', 'type', 'approach', 'important', 'main', 'range',*

*'helpful', 'large', 'difficult', 'available', 'amount', 'useful', 'importance', 'article', 'abstract', 'scale', 'copyright',*

*'real', 'quality', 'inconvenience', 'benefit', 'unavailable', 'term', 'condition', 'interest', 'organization', 'use',*

*'task', 'student', 'professor', 'teacher', 'university']*

lda\_model = LdaMulticore(corpus=corpus, id2word=dictionary, iterations=**500**, num\_topics=1, workers=4, passes=**200**)

Id = 534:  
 0: 0.011\*"**mobile device**" + 0.009\*"**opportunistic network**" + 0.005\*"**cloud computing**" + 0.004\*"information product" + 0.004\*"service updated" + 0.004\*"fault tolerance" + 0.004\*"technology humanity" + 0.004\*"site agreement" + 0.004\*"profit world" + 0.004\*"account management"

Id = 562:

0: 0.017\*"**scheduling algorithm**" + 0.016\*"**cloud computing**" + 0.010\*"**resource management**" + 0.008\*"**satellite image**" + 0.007\*"**service level**" + 0.007\*"**smart city**" + 0.007\*"**cloud service**" + 0.006\*"**time series**" + 0.006\*"**fault tolerant**" + 0.006\*"**genetic algorithm**"

Id = 1146:

0: 0.016\*"**virtual reality**" + 0.007\*"**virtual environment**" + 0.005\*"**sound localization**" + 0.005\*"**sensory substitution**" + 0.005\*"**virtual space**" + 0.005\*"**assistive device**" + 0.004\*"virtual world" + 0.004\*"smith chart" + 0.003\*"fear level" + 0.003\*"sound source"

Id = 841:

0: 0.013\*"**x ray**" + 0.010\*"**electron microscopy**" + 0.007\*"**drug delivery**" + 0.006\*"**composite material**" + 0.005\*"ray diffraction" + 0.004\*"iron oxide" + 0.004\*"sol gel" + 0.004\*"microscopy sem" + 0.003\*"thin film" + 0.003\*"mesoporous silica"

Id = 872:

0: 0.016\*"**neural network**" + 0.009\*"**fractal dimension**" + 0.009\*"**wireless sensor**" + 0.007\*"**unmanned aerial**" + 0.006\*"**image processing**" + 0.006\*"**sensor network**" + 0.005\*"**convolutional neural**" + 0.005\*"**artificial intelligence**" + 0.005\*"**energy consumption**" + 0.005\*"**optic disc**"

Id = 1284:

0: 0.012\*"**natural language**" + 0.009\*"**e learning**" + 0.007\*"**textual complexity**" + 0.007\*"**chat conversation**" + 0.006\*"**social network**" + 0.004\*"polyphonic model" + 0.004\*"learning process" + 0.003\*"artificial intelligence" + 0.003\*"machine learning" + 0.003\*"collaborative learning"