Large-Scale Hierarchical Topic Models

Jay Pujara Department of Computer Science University of Maryland College Park, MD 20742 jay@cs.umd.edu Peter Skomoroch LinkedIn Corporation 2029 Stierlin Ct. Mountain View, CA 94043 pskomoroch@linkedin.com

Abstract

In the past decade, a number of advances in topic modeling have produced sophisticated models that are capable of generating hierarchies of topics. One challenge for these models is scalability: they are incapable of working at the massive scale of millions of documents and hundreds of thousands of terms. We address this challenge with a technique that learns a hierarchy of topics by iteratively applying topic models and processing subtrees of the hierarchy in parallel. This approach has a number of scalability advantages compared to existing techniques, and shows promising results in experiments assessing runtime and human evaluations of quality. We detail extensions to this approach that may further improve hierarchical topic modeling for large-scale applications.

1 Motivation

With massive datasets and corresponding computational resources readily available, the Big Data movement aims to provide deep insights into real-world data. Realizing this goal can require new approaches to well-studied problems. Complex models that, for example, incorporate many dependencies between parameters have alluring results for small datasets and single machines but are difficult to adapt to the Big Data paradigm.

Topic models are an interesting example of this phenomenon. In the last decade, a number of sophisticated techniques have been developed to model collections of text, from Latent Dirichlet Allocation (LDA)[1] through extensions using statistical machinery such as the nested Chinese Restaurant Process [2][3] and Pachinko Allocation[4]. One strength of such approaches is the ability to model topics in a hierarchical fashion. Coarse and fine topics are learned jointly, creating a synergy at multiple levels from shared parameters. However, past experiments have focused on small corpora, such as sampled abstracts from journals with only thousands of documents and terms. Operating on the scale of text collections the size of Wikipedia - millions of documents with millions of terms - is beyond such models using current inference techniques.

While complex models have shown strong analytical results, simpler models that use fewer parameters and produce less structured output are being used at scale. In recent years, many parallel versions of LDA (eg. [5][6][7]) have been developed, capable of summarizing web-scale collections. Although these tools are a boon for analyzing massive amounts of data, they lack the ability to learn a hierarchical representation of topics. This situation highlights a common dilemma in Big Data: whether we can have our cake (models with rich output) and eat it too (operate on massive datasets).

We propose a solution that addresses this dilemma, reusing infrastructure from an existing parallel implementation of LDA in a novel way. Our method provides a scalable mechanism for learning hierarchies from large text collections. We learn top-down hierarchies, first learning topic models at a coarse level, then splitting the corpus into these learned topics and iteratively learning subtopics

in parallel. In experiments on large datasets from Wikipedia and TREC, we show fast training times and favorable human interpretability results, supporting exploratory data analysis at web-scale.

2 Background

Latent Dirichlet Allocation (LDA)[1] is a sophisticated topic model that has served as the foundation for much recent work in topic modeling. The method, defined for k topics and d documents, models topics as distributions over words (β_k), and documents as mixtures of topics (θ_d). For each word in a document d, $w_{d,n}$, a topic $z_{d,n}$ is chosen from the document's topic distribution (θ_d), and a word is chosen from the topic's distribution over words (β_k). The Dirichlet distribution of θ_d and β_k are parameterized by terms α and λ respectively. This generative story is reflected in the joint distribution shown in Equation 1.

$$p(w, z, \theta, \beta | \alpha, \lambda) = \prod_{k} p(\beta_k | \lambda) \prod_{d} p(\theta_d | \alpha) \prod_{n} p(z_{d,n}) p(w_{d,n} | \beta_{z_{d,n}})$$
(1)

Topic models are learned through inference on the generative model described. This inference problem is intractable at scale, and approximate techniques such as Gibbs Sampling or using variational methods are used during modeling. Zhai et al. [7] have argued that variational inference is best suited to learning topic models in large-scale settings, as intra-document dependencies can be minimized through the choice of a variational distribution. Supporting this claim, they've provided an open-source, Apache-licensed implementation for topic modeling called Mr.LDA for Hadoop.

Hierarchical topic models [2] are often in the form of a DAG where nodes correspond to topics at varying levels of granularity. A number of models have been proposed for learning hierarchical topic models, but a common strategy is to choose a *set of topics* for each word in the document, corresponding to a path in the hierarchy. Using such models allows us to understand the relationship between topics at differing levels of the hierarchy, but at a cost. Instead of storing parameters for each term for each of k topics, we must now consider parameters for each *path* in the hierarchy, or k^l parameters for an *l*-level hierarchy with constant branching factor (and potentially more in techniques such as PAMs). Each of these parameters must maintain values for each term in our vocabulary V. The domains where hierarchies might prove most useful - massive, diverse collections using natural language - are likely to have very large vocabularies.

3 Method and Discussion

Algorithm 1 ITERATIVE-TOPIC-HIERARCHY: iteratively learn a topic hierachy

```
Require: Dataset D

Require: Parameters k, LEVELS

Require: LEARN-TOPICS, learn topics from data, as in Mr.LDA : produces \theta_{1:D} and \beta_{1:k}

Require: SPLIT-CORPUS, distribute dataset into k parts using topic-model M

LEARN-NODE(k, D, 0, LEVELS)

function LEARN-NODE(k, D, l, LEVELS)

if l < \text{LEVELS then}

\langle \theta_{1:D}, \beta_{1:k} \rangle = \text{LEARN-TOPICS}(k, D)

D_{1:k} = \text{SPLIT-CORPUS}(\theta_{1:D}, D)

for i = 1 \rightarrow k do

LEARN-NODE(k, D<sub>i</sub>, l + 1, LEVELS)

end for

end if
```

end function

Our approach learns a top-down hierarchy iteratively and is summarized in Algorithm 1. First, we learn a topic model of k topics using the entire corpus. Next, we create k new datasets and allocate each document in the corpus to zero, one, or more of the k datasets using method SPLIT-CORPUS and the θ values learned by the topic model. Finally, we learn k new topic models using the k

datasets constructed in the previous step. Each of these topic models corresponds to subtopics of one of the original k topics. This procedure can be applied iteratively, generating further levels of the hierarchy. Since our method doesn't add dependencies from subtopics to their parents (or siblings), the process is easily parallelizable by launching each learning task independently.

The consequence of these decisions allows us to leverage massive datasets while escaping some of the limitations of current hierarchical topic modeling. Existing models attempt to learn parameters *jointly* across levels of the hierarchy, while our approach learns at a single level, implicitly conditioning on ancestors. Instead of learning parameters for each *path* in the hierarchy, we learn parameters for each *node*, removing dependencies between nodes and allowing learning to take place in parallel. Finally, our method operates in a coarse-to-fine fashion, first learning a flat topic model and refining each topic, using information from the coarse model to make decisions about these refinements through its allocation of documents to subtasks.

4 Implementation Details

Our implementation uses the Mr.LDA package to learn the topic model (LEARN-TOPICS in Algorithm 1). We make customizations to this package to allow easier parallelization of learning and add support three models providing the functionality of SPLIT-CORPUS:

- SPLIT-SINGLE chooses the highest probability topic for each document d using θ_d and allocates the document to the corresponding topic: $[D_k = \{d : mode(\theta_d) = k\}]$
- SPLIT-MULTIPLE creates c copies of the document, and proportionally allocates those documents according to the probabilities in θ_d : $[D_k = \{d \times \text{cnt} : \text{cnt} = \text{round}(c * \theta_d[k])\}]$
- SPLIT-SELECT applies a threshold, t, on the entropy (H) of θ_d and then allocates documents below the threshold to the highest probability topic (as in SPLIT-SINGLE): $[D_k = \{d : \text{mode}(\theta_d) = k \land H(\theta_d) < t\}]$

SPLIT-SINGLE provides a very simple method to split the corpus, and has the added advantage that each document is used exactly once in each level of the hierarchy. SPLIT-MULTIPLE embodies the tenet of LDA topic modeling that each document is a mixture of topics by creating multiple copies of each document. A drawback of this approach is that the number of documents grows with each level, c^l , although this can be addressed with horizontal scaling. Another drawback is that, due to the allocation of less relevant documents to a topic, incoherent subtopics may be learned. Finally, SPLIT-SELECT chooses only documents that strongly map to a particular topic by using a threshold based on the entropy of the θ_d distribution. This allows us to cull irrelevant documents, but possibly at the cost of losing specialized subtopics that occur rarely in the dataset.

5 Datasets and Results

Our method was applied to two large-scale document collections - TREC and Wikipedia. The TREC collection consists of 473K documents from the Financial Times and LA Times[8]. Tokens were processed by stemming and removing terms that occur less than 20 times, yielding a vocabulary of 60K. The Wikipedia dataset consists of 3M documents with a vocabulary of 1.8M after removing tokens that either occurred in more than 90% of documents or less than 20 times. Illustrative results are also presented for LinkedIn member profiles in the Appendix.

For both of these collections, we learned a 2-level topic hierarchy on a modest Hadoop cluster. For the TREC corpus, we used a branching factor of 5 (5 topics at the root, 5 subtopics for each topic). For each learning phase, we used 40 mappers and 20 reducers. The root topic model ran for 22 iterations of variational inference with an average iteration time of 360s, after which the model converged (log likelihood changed by less than a factor of 10^{-6}). Subtopics from the root were run for 20 iterations, with an average iteration time of 230s using SPLIT-SINGLE corpus allocation. The total time for learning the full hierarchy was approximately 3.5 hours. For the larger Wikipedia corpus, we used a branching factor of 10 (10 root topics with 10 subtopics each). The root topic model converged after 41 iterations, with an average iteration time of 230os. Subtopics from the root were run for 20 iterations, with an average iteration time of 700s using SPLIT-SINGLE

SINGLE	japan	los	bosnia	america	hong
	korea	metro	serb	panama	kong
	moscow	angel	london	colombia	li
	nuclear	orang	attack	prevent	provinci
	russia	san	kill	latin	comrad
MULTIPLE	report	mr	column	94	parti
	94	govern	counti	report	govern
	forc	time	citi	1994	94
	govern	industri	who	english	report
	militari	financi	part	peopl	state
SELECT	industri	traffick	tass	list	evolutionari
	london	santa	africa	noriega	comrad
	moslem	mexico	itar	won	idea
	ft	park	sov	seat	chuch
	product	beach	herzegovina	known	procuratori

Table 1: Three sets of subtopics of topic 4 from the TREC corpus (brief/investig/korea/beij) generated by different document-allocation techniques. (Full TREC hierarchy in Appendix)

Measure	SINGLE	MULTIPLE	SELECT
Total Documents	472524	2362303	333693
Iteration time (s)	230	405	77
Intruders Found (%)	27.4	13.0	25.5

Table 2: Evaluation of scalability and quality of document-allocation techniques

corpus allocation. The total time for learning the full hierarchy was approximately 30 hours. The full hierarchy, with the top 5 IDF-normalized terms for each topic, is shown in the Appendix.

We proposed three methods for splitting the dataset when allocating documents to subtrees, and sample output for each is shown in Table 1. SPLIT-MULTIPLE was run with c = 5, so that 5 copies of each document were generated and allocated among the 5 possible subtopics. SPLIT-SELECT was run with t = .72, where documents where the entropy of θ_d was less than .72 were assigned to the most-probable topic in the multinomial and all others were ignored.

Speed and quality for these methods are assessed in Table 2. Adding multiple copies of documents increased running time by 75%, while only choosing low-entropy documents decreased running time by 66%. A growing trend in topic models is to focus on interpretablity rather than model log likelihood [9], using *word intrusion* as a metric. Humans are presented with the top terms for a topic as well as an intruder - a term that has low probability for the current topic but high probability for some other topic. We conducted such an evaluation with six participants, using the 10 top terms in each topic as well as an intruder term, so that randomly choosing a term would have a 9.1% success rate. The intruder term was among the top 100 terms for another topic in the same subtree, but ranked between 250-500 for the chosen topic. Terms in each list were presented in random order, and the term lists were randomized across methods. The results suggest that the MULTIPLE (15.9%) strategy lags behind both SINGLE (27.4%) and SELECT (25.5%) which have similar results.

6 Conclusion and Future Work

The hierarchical topic modeling approach we've presented fulfills two goals outlined - rich output and scalability. The hierarchies learned are interpretable, both qualitatively and based on a wordintrusion evaluation. The method is scalable, operating on millions of documents and terms and producing results on the scale of hours and days. One aspect of our approach is allocating documents to subtasks learning different branches of the hierarchy. Our evaluation suggests that single-allocation or a selective-allocation both perform well, with a small quality-scalability tradeoff. In future work, we hope to perform more extensive evaluation, including a comparison of these techniques to prior models trained on a smaller sample of data. We are also interested in implementing a parallel variational inference algorithm for hierarchical LDA that uses sketches or hashing to reduce the number of parameters.

References

- [1] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.
- [2] David M. Blei, Thomas L. Griffiths, Michael I. Jordan, and Joshua B. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. In *NIPS*, 2003.
- [3] Chong Wang and David M. Blei. Variational inference for the nested chinese restaurant process. In NIPS, 2009.
- [4] Wei Li and Andrew Mccallum. Pachinko Allocation: Scalable Mixture Models of Topic Correlations. Journal of Machine Learning Research, 2008.
- [5] Ramesh Nallapati, William W. Cohen, and John D. Lafferty. Parallelized variational em for latent dirichlet allocation: An experimental evaluation of speed and scalability. In *ICDM Work*shops, 2007.
- [6] Feng Yan, Ningyi Xu, and Yuan Qi. Parallel inference for latent dirichlet allocation on graphics processing units. In NIPS, 2009.
- [7] Ke Zhai, Jordan Boyd-Graber, Nima Asadi, and Mohamad Alkhouja. Mr. LDA: A flexible large scale topic modeling package using variational inference in mapreduce. In *WWW*, 2012.
- [8] National Institute of Standards and Technology. Special db 22 and nist tree document database. http://www.nist.gov/srd/text.cfm.
- [9] Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. Reading tea leaves: How humans interpret topic models. In *NIPS*, 2009.

7 Appendix

The customizations to Mr.LDA we detail are freely available at https://github.com/puuj/Mr.LDA

Terms shown in the tables below are normalized using a formula similar to TF-IDF. The probability of the term is used in lieu of term frequency, and inverse *topic* frequency is used to decrease the weight of terms that appear in multiple topics at the same level of the hierarchy. In the output shown, the top 500 terms from each topic were used for this normalization.

To save space, hierarchies are shown as tables in the Appendix. The first row of each table corresponds to a topic at the root level. Subsequent entries in each column are subtopics: children of the topic seen in the first row.

topic 0	topic 1	topic 2	topic 3	topic 4
design	dollar	game	al	brief
materi	share	team	arab	investig
use	uk	sport	palestinian	korea
qualiti	profit	season	israel	beij
space	stock	photo	isra	letter
subtopic 00	subtopic 1 ₀	subtopic 20	subtopic 30	subtopic 4 ₀
comput	appoint	shop	resourc	japan
softwar	airlin	museum	task	korea
electron	sir	restaur	materi	moscow
machin	labour	buy	labor	nuclear
user	court	fashion	properti	russia
subtopic 01	subtopic 1_1	subtopic 2 ₁	subtopic 31	subtopic 4 ₁
music	pension	inning	uk	los
artist	risk	pitch	pound	metro
news	life	goal	93	angel
garden	save	touchdown	gatt	orang
british	mortgag	led	round	san
subtopic 0 ₂	subtopic 1_2	subtopic 2 ₂	subtopic 3 ₂	subtopic 4 ₂
subtopic 0 ₂ beij	index	subtopic 2 ₂ russian	battalion	subtopic 4 ₂ bosnia
· · -	· · ·			· -
beij	index	russian	battalion	bosnia
beij farmer	index 00	russian concert	battalion artilleri	bosnia serb
beij farmer daili fbis provinc	index 00 se qtr 90	russian concert moscow type radio	battalion artilleri ukrainian enemi zhirinovski	bosnia serb london attack kill
beij farmer daili fbis provinc subtopic 0 ₃	index 00 se qtr	russian concert moscow type radio subtopic 2 ₃	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃	bosnia serb london attack kill subtopic 4 ₃
beij farmer daili fbis provinc	index 00 se qtr 90 subtopic 1 ₃ pre	russian concert moscow type radio	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat	bosnia serb london attack kill
beij farmer daili fbis provinc subtopic 0 ₃	index 00 se qtr 90 subtopic 1 ₃ pre interim	russian concert moscow type radio subtopic 2 ₃ driver feet	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo	bosnia serb london attack kill subtopic 4 ₃ america panama
beij farmer daili fbis provinc subtopic 0 ₃ ecolog command weapon	index 00 se qtr 90 subtopic 1 ₃ pre interim fin	russian concert moscow type radio subtopic 2 ₃ driver	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat	bosnia serb london attack kill subtopic 4 ₃ america
beij farmer daili fbis provinc subtopic 0 ₃ ecolog command weapon coordin	index 00 se qtr 90 subtopic 1_3 pre interim fin 5m	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent
beij farmer daili fbis provinc subtopic 0 ₃ ecolog command weapon coordin combat	index 00 se qtr 90 subtopic 1_3 pre interim fin 5m 3m	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin
beij farmer daili fbis provinc $subtopic 0_3$ ecolog command weapon coordin combat $subtopic 0_4$	index 00 se qtr 90 subtopic 1 ₃ pre interim fin 5m 3m subtopic 1 ₄	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride subtopic 2 ₄	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo subtopic 3 ₄	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin subtopic 4 ₄
beij farmer daili fbis provinc $subtopic 0_3$ ecolog command weapon coordin combat $subtopic 0_4$ communiti	index 00 se qtr 90 subtopic 1 ₃ pre interim fin 5m 3m subtopic 1 ₄ percent	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride subtopic 2 ₄ attorney	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo subtopic 3 ₄ bosnia	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin subtopic 4 ₄ hong
beij farmer daili fbis provinc $subtopic 0_3$ ecolog command weapon coordin combat $subtopic 0_4$ communiti district	index 00 se qtr 90 subtopic 1 ₃ pre interim fin 5m 3m subtopic 1 ₄ percent china	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride subtopic 2 ₄ attorney law	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo subtopic 3 ₄ bosnia muslim	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin subtopic 4 ₄ hong kong
beij farmer daili fbis provinc subtopic 0 ₃ ecolog command weapon coordin combat subtopic 0 ₄ communiti district resid	index 00 se qtr 90 subtopic 1 ₃ pre interim fin 5m 3m subtopic 1 ₄ percent china nuclear	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride subtopic 2 ₄ attorney law council	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo subtopic 3 ₄ bosnia muslim polic	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin subtopic 4 ₄ hong kong li
beij farmer daili fbis provinc $subtopic 0_3$ ecolog command weapon coordin combat $subtopic 0_4$ communiti district	index 00 se qtr 90 subtopic 1 ₃ pre interim fin 5m 3m subtopic 1 ₄ percent china	russian concert moscow type radio subtopic 2 ₃ driver feet accid grade ride subtopic 2 ₄ attorney law	battalion artilleri ukrainian enemi zhirinovski subtopic 3 ₃ arafat plo jordan jan cairo subtopic 3 ₄ bosnia muslim	bosnia serb london attack kill subtopic 4 ₃ america panama colombia prevent latin subtopic 4 ₄ hong kong

Table 3: Topic Hierarchy with branching factor 5 for TREC-60K corpus using SPLIT-SINGLE

windsrecordsopicesopicespindroue64cablesleigueelectionannotfesticalfoodconstantfried.lakeoptbishopfoodconstantinstantannotalfoodconstantfried.lakeoptsublepic 50,sublepic 51,sublepic 51,sub	topic 0	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6	topic 7	topic 8	topic 9
physical cannotrecorded (statu)enemy (statu)epipoles (statu)minute (statu)enemy (statu)minute (statu)enemy (statu)										
CommonInstituteTrendInke*44HoriCurpinstituteindividualmusicalfisharmedvortantkmhorizeamposhipprofessorabbrogit Cosubopit Cosubopit Cosubopit Cosubopit Cosubopit Cosubopit Coprofessorwikadatacassvortantkmerraerrarectanotmovieusesfinalesuccesvortantkmerraerrarectanotsakscolsfinalesuccesvortantsubopit Cosubopit Cosubopit Cosubopit Cosakscolsfinalesuccessubopit Cosubopit Cosu										
imital imital										
individualmoscaloriginorigingenerationkindgenerationkindgenerationkindgenerationkindgenerationkindgenerationkindgenerationkindgenerationkindgeneration <thg< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></thg<>										
subtopie 0) Iordsubtopie 10 screessubtopie 11 screessubtopie 10 s										
led secres binds gas wwe fa eighteens swedens picher means movie uses fernale scholan tag spanish sixty swedish asservid cannot ass tal ast anged trag spanish sixty swedish asservid cannot subtopic 01 subtopic 11 subtopic 21 subtopic 22 subtopic 23							1 7			1
wife movieuses usesmean fermalecone cane asternicmean tasternicmean cane canonmean canonmea		subtopic 10		subtopic 30	subtopic 40					
moviesusesfermilescholarstagspanishstrystrysweldshasteroidcannotasksnilemaleworldsriplecl6denaratpichedfilteroferesathagie Crysathogie Crysubtopie	lord	scenes	birds	gas	wwe	la	eighteens	sweden	pitcher	means
askstellstailargedravbrzil4'Mukandstor,drugsubtop(C1)subtop(wife	takes	males	cases	wwf	del	eighteen	anne	era	treatment
sectoridenmakeworldsitripleeff6f6fdenmarkpicheditherforesettor(c) <td>movie</td> <td>uses</td> <td>female</td> <td>scholars</td> <td>tag</td> <td>spanish</td> <td>sixty</td> <td>swedish</td> <td>asteroid</td> <td>cannot</td>	movie	uses	female	scholars	tag	spanish	sixty	swedish	asteroid	cannot
search dancers subtopie 21 dancers subtopie 21 served subtopie 22 served subtopie 21 served subtopie 22 served subtopie 22 served subtopie 22 served subtopie 23 subtopie 23 subtopie 23 subtopie 23 subtopie 23 subtopie 23 subtopie 23 subtopie 24 subtopie 25 subtopie 2	asks	tells	tail	argued	raw	brazil	4	husband	sox	drug
search databasedancing teaservedfilm singhhollywoodstorm rolesstorm roleselementary valleyubare pagepain songpoint pagekin republicaemain pagechinesestylelovefestival famingstructurepostmusicalkickerwriter newspaperssubtopic 02 gamesubtopic 13 subtopic 14subtopic 13 subtopic 13subtopic 13 subtopic 13 subtopic 13subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 14subtopic 13 subtopic 24 subtopic 23 subtopic 23 subtopic 24 subtopic 25 subtopic 26 subtopic 26 <b< td=""><td>scene</td><td>idea</td><td>male</td><td>worlds</td><td>triple</td><td>el</td><td>6</td><td>denmark</td><td>pitched</td><td>therefore</td></b<>	scene	idea	male	worlds	triple	el	6	denmark	pitched	therefore
search databasedancing teaservedfilm singhhollywoodstorm rolesstorm roleselementary valleyubare pagepain songpoint pagekin republicaemain pagechinesestylelovefestival famingstructurepostmusicalkickerwriter newspaperssubtopic 02 gamesubtopic 13 subtopic 14subtopic 13 subtopic 13subtopic 13 subtopic 13 subtopic 13subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 13 subtopic 14subtopic 13 subtopic 24 subtopic 23 subtopic 23 subtopic 24 subtopic 24 	subtopic 01	subtopic 11	subtopic 21	subtopic 31	subtopic 41	subtopic 51	subtopic 61	subtopic 71	subtopic 81	subtopic 91
databasedancingtendsinghrolestropicalvalley'openkicktelevisionpagerepertivestylelovefastraddanagepostmusicalkickerwriterclientdancerdancerfilmingstructurenearbysongdbnewspacersgungamewatelibandpakfilmingstructurenearbysgrbopic 7_2subtopic 8_2subtopic 9_2gungameconcaretewangggittarlackscolroomsseatsviceplacesfindieconcaretewangggittarlackscolroomsseatsvicesubtopic 02subtopic 13subtopic 23subtopic 24subtopic 34subtopic 3										
enailbollywordchinesemsicstardfluringpagerepublicatpipsysongpoliskgallorgathopic 2;subtopic 4;subtopic 4;s										
cheftetdancerdishantfillingstructireneartysongqbnewspapersgunggunggungsubtopic 32subtopic 42subtopic 42 <td></td>										
subtopic 02 gan subtopic 12 gan subtopic 22 wate subtopic 32 band subtopic 42 band subtopic 52 band subtopic 52 wice subtopic 53 wite subtopic 54 wice subtopic 54 wite subtopic 55 wite subtopic 55 wite <td></td> <td></td> <td></td> <td></td> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td>						0				
gain planetsgaine consumptionwate formula albumin albumpeak albud mrckmind windmeand moronsmeand seatsbyspacernft lunchindic lubelsconcrete formulawang planetsgameplay columbiacolumbia sationroomsseating consumptionlunchlabelsfarmhan playstationsubtopic 53 subtopic 53 subtopic 53 subtopic 53 subtopic 53 subtopic 54subtopic 53 subtopic 54 subtopic 54subtopic 53 subtopic 54 subtopic 54subtopic 53 subtopic 54 subtopic 54subtopic 54 subtopic 54 subtopic 54subtopic 54 subtopic 55subtopic 54 subtopic 55subtopic 54 subtopic 55subtopic 54 subtopic 55subtopic 57 subtopic 55subtopic 57 subtopic 55subtopic 57 subtopic 54subtopic 56 subtopic 57 subtopic 57subtopic 57 subtopic 57 subtopic 57subtopic 57 subtopic 57 subtopic 57subtopic 57 subtopic 57 subtopic 57subtopic 57 subtopi										
planets spacecraft hebrewelectronic indicconsumption concreteformula guitaralbam guitarrock lawwind rood roodrooms construction illinoislauch hebrewfullength amitoticensistonslia guitarguitar guitarlakes rodfeatures roodseating seatingdemocrat empire jerseysubtopic 0.3 subtopic 0.3 actionssubtopic 3.3 subtopic 3.3subtopic 3.3 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4 subtopic 3.4 subtopic 3.4 subtopic 3.4 subtopic 3.4subtopic 3.4 subtopic 3.4										
spacecraft lebrewindic fullengelconcretevang gamelyacolum playstationcolum summitcolum plantsroof railwaysconstructionillinoislauchlabelsfarmhanplaystationsummitplantstopempirejerseyabtopic 03subtopic 03subtopic 03subtopic 03subtopic 03subtopic 03subtopic 03subtopic 03subtopic 03subtopic 04actionsscalefloweringmirishwindowsgermanymikshipoic 04subtopic 04subtopic 04consciousnessfrvvfloweringsubtopic 04subtopic 04subtopic 04subtopic 04subtopic 04subtopic 04subtopic 04guitarinterviewproteinconomyemperorfidesubtopic 04subtopic 04subtopic 04subtopic 04subtopic 04anouncedaskdeproteinsfirinliconpanyscommunicationsbutalionridersparlianentaryautoristingreallygeneastrinaliconpanyscommunicationsbutalionridersparlianentaryanouncedaskdeproteinsfirinafirinaliconpanyscommunicationsbutalionriderssubtopic 05,subtopic 05,subtopic 15,subtopic 15,subtopic 15,subtopic 15,subtopic 15,subtopic 15,subtopic 15,subtopic 15,subtopic 15,anouncedfedurencerearreinrearreinre										
herewfullengthemissionsliugameplaycolumbiarelawayfeaturesseatingdemocratlauchlabelsfarmhanplaystionsubtopic 73subtopic 74subtopic 75subtopic 75subtopi										
landchlandpinystationsummitppinstopempinejerseysubtopic 3subtopic 3subtopic 3subtopic 4subtopic 3subtopic 4subtopic 3subtopic 4subtopic 4su										
subtopic 03 disorder subtopic 13 regen subtopic 13 subtopic 23 subtopic 33 subtopic 34 subtopic 35 subtopic 34 subtopic 35 subtopic 35 subtop		6					-		0	
descrider actionsscale flowersirish minuwindowsindia indiafuselage messship shipwicketsscientific ocontionsconstrive self revolutionirini moistsingapore irawer wer wer irapoland alitudealitude fl fl subtopic 64.shipoi fb subtopic 64.shipoi subtopic 74.subtopic 84. subtopic 84.subtopic 85. subtopic 95.subtopic 85. subtopic 85.subtopic 85					1 2		1			
actions cognitive selfscale irrainan mothforwer irrainan mothumiform irrainand mothuser irrainand datagermany hiter hiteracy meter the hmsships athight fithships mass hiteracy fithships mass hiteracy fithships mass hiteracy fithships mass hiteracy fithships mass hiteracy hiteracy fithships hiteracy hiteracy hiteracy fithships hiteracy hi	subtopic 03	subtopic 13	subtopic 23	subtopic 33	subtopic 43	subtopic 53	subtopic 63	subtopic 73	subtopic 83	subtopic 93
cognitive selfimmain revolutionmoist ireland irelandsingapore userweb ireland"poland' atteracy prefecturealtitude kgflet man batsmanmanchester physics admiraleconomics physics anvionmentalsubtopic 04 subtopic 04 guitarsubtopic 14 interviewsubtopic 12 asubtopic 14 subtopic 14subtopic 04 subtopic 34 asiasubtopic 04 subtopic 04 asiasubtopic 04 subtopic 04 subtopic 05 subtopic 05 su	disorder	ragam	subtropical	irish	windows	india	fuselage	ship	wickets	scientific
selfrevolutionmothirelanddataİteracyfthmscricketerptysicsconsciousnessfivfibroitsubtopic 24subtopic 34subtopic 54subtopic 54subtopic 54subtopic 64regimenttaglaboragularinterviewproteineconomyemperorfideinternetregimenttaglaboraadvertisingreallygeneafricanlicompanyscommunicationsbatalionriderspaliamentaryatomsaskingdaarmenianchineseliverymicrosoftmedalmvestercabiretssubtopic 05subtopic 05subtopic 75subtopic	actions	scale	flowers	uniform	users	germany	mk	ships	firstclass	phd
consciousnessfwvfloweringirauserprefecturekgadmiralbatmanenvironmentalsubtopic 0_4 subtopic 1_4 subtopic 2_4 subtopic 3_4 subtopic 5_4 subtopic 6_4 subtopic 6_4 subtopic 6_4 subtopic 8_4 subtopic 9_4 v wantedcellasiaprincebrandwirelessbrigadepokersatsadvertisingreallygeneafricanlicompanyscommunicationsbattalionridersanouncedaskedproteinstaiwanliuwentvideoawardedwrestlercoabinetsubtopic 0_5 subtopic 2_5 su	cognitive	iranian	moist	singapore	web	poland	altitude	fleet	manchester	economics
subtopic 04 guitarsubtopic 14 interviewsubtopic 24 proteinsubtopic 34 economysubtopic 44 emperorsubtopic 54 ridesubtopic 64 internetsubtopic 74 interessubtopic 75 interessubtopic 75 <br< td=""><td>self</td><td>revolution</td><td>moth</td><td>ireland</td><td>data</td><td>literacy</td><td>ft</td><td>hms</td><td>cricketer</td><td>physics</td></br<>	self	revolution	moth	ireland	data	literacy	ft	hms	cricketer	physics
subtopic 04 guitarsubtopic 14 interviewsubtopic 24 proteinsubtopic 24 economy economy asiasubtopic 54 emperorsubtopic 54 indesubtopic 74 inde brandsubtopic 74 intervetsubtopic 74 inde brandsubtopic 74 intervetsubtopic 74 intervet<	consciousness	fwv	flowering	ira	user	prefecture	kg	admiral	batsman	environmental
guitar tvinterview wardedprotein taiaeconomy asiaemperor prince brandinternet brand wardedregiment brigade wardedtag policionlabour seassadvertising announcedraelly askedgene proteinsafrican taiwanliu liucompanys communicationscommunications awardedbutation awardedriders parliamentary cabinetsubtopic 05 subtopic 15 subtopic 15subtopic 15 subtopic 15subtopic 35 subtopic 35subtopic 45 subtopic 45 subtopic 55subtopic 65 subtopic 65 subtopic 75subtopic 75 subtopic 75subtopic 35 subtopic 35subtopic 35 subtopic 35subtopic 35 subtopic 35subtopic 35 subtopic 35subtopic 65 subtopic 65subtopic 75 subtopic 75subtopic 35 subtopic 35subtopic 35 subtopic 36subtopic 35 subtopic 36subtopic 35 subtopic 36subtopic 35 subtopic 36subtopic 35 subtopic 36subtopic 35 subtopic 56subtopic 65 subtopic 66subtopic 76 subtopic 76subtopic 86 subtopic 86 subtopic 96subtopic 37 subtopic 37 subtopic 37subtopic 37 subtopic 37 subtopic 37subtopic 37 subtopic 36subtopic 37 subtopic 37 subtopic 37subtopic 37 subtopic 37 subto	subtopic 04	subtopic 14	subtopic 24	subtopic 34	subtopic 44	subtopic 54		subtopic 74	subtopic 84	subtopic 94
tvwantedcellasiaprincebrandwirelessbigadepokerseatsadvertisingreallygeneafricanlicompanyscommunicationsbattalionridespoteinsawardedwrestlerconstituencystabtopicsakedproteinstaiwanlinliverymicrosoftmedalmreadmrasconstituencysubtopic <td></td>										
advertising announcedreally askedgene proteinsafrican taiwani liucommunications wentbatalion videobatalion wardedriders calinetparliamentary calinetsubtopic 05 subtopic 15 anglesubtopic 15 subtopic 16subtopic 35 subtopic 35subtopic 45 subtopic 55subtopic 55 subtopic 55subtopic 75 subtopic 75subtopic 95 subtopic 95 communitessubtopic 95 subtopic 95 communitessubtopic 95 subtopic 95subtopic 95 subtopic 95subtopic 95 subtopic 95subtopic 96 subtopic 96subtopic 96 subtopic 96subtopic 96 subtopic 96subtopic 96 subtopic 96subtopic 96 subtopic 96subtopic 96 subtopic 97 subtopic 98 season season season season season season season season season seas	tv									
announced stationsaskedproteinstaiwanliuwerrvideoawardedwwerstlercabiner cabinersubtopic 05 subtopic 05 anglesubtopic 15 signalsubtopic 25 tambonsubtopic 35 divisionssubtopic 35 carssubtopic 55 carssubtopic 75 chinasubtopic 75 taxessubtopic 75 olympicssubtopic 75 constituencysubtopic 75 constituency	advertising								1	
stationssayingdnaarmenianchineseliverymicrosoftmedalnwaconstituencysubtopic 05subtopic 15subtopic 25subtopic 25subtopic 35subtopic 57subtopic 75subtopic 75subtopic 85subtopic 96anglesignaltambondivisionscarschinataxesreferolympichongconnectedfrequencyamphoebattalionscontestantschinasrevenuesurnameolympichongcurvetransmittervilagesoffensiveroundhongratesarabicsilverhongindexoperatedtownpositionscontestantschineserevenuesurnameolympicshongletsiteeruptionarmouredaircraftoilpaycodexmetresneedssubtopic 06subtopic 16subtopic 26subtopic 46subtopic 56subtopic 76subtopic 76subtopic 96particlestaredrailwaygreecegetspayheatcancersantamclarenbudgetacidproductionsfoundedserbiadecidescorridorbatteryportugalflsamesxelectronbatopic 17subtopic 27subtopic 37subtopic 75subtopic 75subtopic 75subtopic 75subtopic 75subtopic 17subtopic 17subtopic 29subtopic 37subtopic 37subtopic 38serbiacorridor<		-								
subtopic 05 anglesubtopic 15 signalsubtopic 25 tambonsubtopic 35 divisionssubtopic 45 carssubtopic 55 takessubtopic 65 takessubtopic 75 refersubtopic 85 olympicssubtopic 95 communitiescurve indexfrequency operatedamphoe townpositionscontestants ontestantschinae contestantstakes takesrefer takesolympicscommunities olympicssubtopic 06 battalionssubtopic 26 sitesubtopic 26 subtopic 26subtopic 36 subtopic 36 subtopic 36subtopic 46 subtopic 46 subtopic 46subtopic 76 subtopic 76 subtopic 76 subtopic 76subtopic 76 subtopic 76 subtopic 76 subtopic 76 subtopic 76 subtopic 77 subtopic 76 subtopic 76subtopic 76 subtopic 76 subtopic 76 subtopic 76 subtopic 77 subtopic 77 subtopic 76 subtopic 76 subtopic 77 subtopic 76 subtopic 76 subtopic 76 subtopic 77 subtopic 77 subtopic 76 subtopic 76 subtopic 77 subtopic 777 subtopic										
anglesignaltambondivisionscarschinataxesreferolympicscomunitiesconnectedfrequencyamphoebattalionscontestantschineserevenuesurnameolympichongcurvetransmittervillagesoffensiveroundhongratesarabicsilverkongindexoperatedtownpositionscontestantkonghealthmanuscriptbronzeprovidingletsiteeruptionarmouredaircraftoilpaycodexmetresneedssubtopic 06subtopic 16subtopic 26subtopic 36subtopic 56subtopic 66subtopic 76subtopic 86subtopic 96particlestaredrailwaygreecegetspaykayheatcathedralferraribillioncancerseasonservicesberlintriestumpikerequirementspapalschunacheramericansacidproductionsfoundedserbiadecidescorridorbattalionsfilsamesexsubtopic 07subtopic 17subtopic 27subtopic 37subtopic 47subtopic 57subtopic 67subtopic 77subtopic 87subtopic 97chinesepolandstarlatrialtrialterminalairfiedchristiansfifabishopverbsdistrictsyndromecharlesarrestedairlinemasygodsmidfieldercommand										
connectedfrequencyamphoebattalionscontestantschineserevenuesumameolympichongcurvetransmittervillagesoffensiveroundhongratesarabicarabicsilverkongindexoperatedtownpositionscontestantkonghealthmanuscriptbronzeprovidingletsiteeruptionarmouredaircraftoilpaycodexmetresneedssubtopic 06subtopic 16subtopic 36subtopic 46subtopic 56subtopic 67subtopic 76subtopic 76subtopic 76subtopic 77subtopic 77subtopic 77subtopic 77subtopic 77subtopic 77subtopic 77subtopic 77subtopic 77subtopic 78bislopeelectronsproductionsfoundedstarlatrialterminalairfieldchiristiansfifabislopic 97chinesepolandstarlatrialtariniairfieldchiristiansfifabislopic 97chinesepolandstarlatrialtariniairfieldchiristiansfifabislopic 97oursedistrictsyndromecharlesarrestedairlinemassignedpraygodsmidfeldercommanderoursevorbogic 18subtopic 28subtopic 28subtopic 58subtopic 58subtopic 58subtopic 77subtopic 77subtopic 77subtopic 97chinesepland <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>										
curvetransmittervillagesoffensiveroundhongratesarabicsilversilverkongindexoperatedtwnpositionscontestantkonghealthmanuscriptbronzeprovidingletsiteeruptionarmouredaircraftoilpaycodexmetresneedssubtopic 06subtopic 16subtopic 26subtopic 36subtopic 46subtopic 56subtopic 66subtopic 76subtopic 86subtopic 96particlestarredrailwaygreecegetsparkmaykheatcathedralferraribillioncancerseasonservicesberlintriesturpikerequirementspapalschumachersubtopic 36electroncastbusinessoccupationaskstollgreatersantamclarenbudgetacidproductionsfoundedserbiadecidescorridorbatteryportugalflsamesexelectronsubtopic 77subtopic 27subtopic 37subtopic 37subtopic 57subtopic 67subtopic 77subtopic 87subtopic 97chinesepoladstarlatrialarrestedairlinessagadronsgodsmidfeldercommanderverbsdistrictsyndromecharlesarrestedairlinemissionsgodsmidfeldercommandersubtopic 08subtopic 18subtopic 28subtopic 38subtopic 55 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>										
index letoperated sitetown eruptionpositions armouredcontestant aircraftkong oilhealth paymanuscript codexbronze metresproviding needssubtopic 06 subtopic 06 subtopic 06 cancersubtopic 16 staredsubtopic 26 railwaysubtopic 26 getssubtopic 56 getssubtopic 66 subtopic 66subtopic 76 subtopic 76 subtopic 76subtopic 76 subtopic 76subtopic 76 subtopic 76subtopic 76 subtopic 76subtopic 76 subtopic 76subtopic 76 subtopic 77subtopic 76 subtopic 77subtopic 77 subtopic 77subtopic 78 subtopic 77subtopic 78 subtopic 78subtopic 78 subtopic 797subtopic 78 subtopic 78subtopic 78 subtopic 797subtopic 78 subtopic 78subtopic 78 subtopic 78subtopi										
letsiteeruptionarmouredairreaftoilpaycodexmetresneedssubtopic 06subtopic 16subtopic 26subtopic 36subtopic 56subtopic 56subtopic 66subtopic 76subtopic 76subtopic 86subtopic 96particlestarredrailwaygreecegetsparkwayheatcathedralferraribillioncancerseasonservicesberlintriestumpikerequirementspapalschumacheramericanselectroncastbusinessoccupationaskstollgreatersantamclarenbudgetelectronsbroadwaylaunchedhungaryleavesbypassspeedsvenicehamiltonincreasedsubtopic 7subtopic 7subtopic 7subtopic 37subtopic 37subtopic 57subtopic 7subtopic 7subtopic 87subtopic 97chinesepolandstarlatrialtrialairfieldchristiansfifabishopverb<										
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $										
particlestarredrailwaygreecegetsparkwayheatcathedralferraribillioncancerseasonservicesberlintriestumpikerequirementspapalschunacheramericanselectroncastbusinessoccupationakstollgreatersantamclarenbudgetelectronsbroadwaylaunchedhungaryleavesbypassspeedsvenicehamiltonincreasedsubtopic 07subtopic 17subtopic 37subtopic 37subtopic 47subtopic 57subtopic 67subtopic 7subtopic 57subtopic 97chinesepolandstarlatrialterminalairfieldchristiansfifabishopverbsdistrictsyndromecharlesarrestedairlinessquadronsgodsgodsserieseminarygenderadministrativemusclepuertotrianairlinemissionsgospeldefendernavygenderadministrativemusclepuertotrianairlinemissionsgospeldefendernavygenderadministrativemusclepuertotrianairlinessubtopic 58subtopic 98subtopic 98subtopic 98subtopic 98subtopic 98subtopic 98seminarygenderadministrativemusclepuertotrianairlinessubtopic 58subtopic 98subtopic 98subtopic 98subtopic 98subtopic			1							
LancerseasonservicesberlinTriesturnpiketrequirementspapalschumacheramericanselectroncastbusinessoccupationaskstollgreatersantamclarenbudgetacidproductionsfoundedserbiadecidescorridorbatteryportugalf1samesexelectronsbroadwaylaunchedhungaryleavesbypassspeedsvenicehamiltonincreasedsubtopic 07subtopic 17subtopic 27subtopic 37subtopic 47subtopic 57subtopic 67subtopic 77subtopic 87subtopic 97chinesepolandstarlatrialterminalairfieldchristiansfffabishopverbsdistrictsyndromecharlesarrestedairlinessquadronsgodsmidfieldercommandernounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrialairlinemissionsgooseldefendernavysubtopic 08subtopic 18subtopic 28subtopic 38subtopic 48subtopic 58subtopic 58subtopic 58subtopic 78subtopic 78subtopic 89subtopic 99superlistingfflyingchanelsleaguev8tibetanncaateamssuperlistingfflyingchanelsleaguev										
electroncastbusinessoccupationaskstollgreatersantamclarenbudgetacidproductionsfoundedserbiadecidescorridorbatteryportugalf1samesexelectronbroadwaylaunchedhungaryleavesbypassspeedsvenicehamiltonincreasedsubtopic 07subtopic 17subtopic 27subtopic 37subtopic 47subtopic 57subtopic 67subtopic 77subtopic 87subtopic 97chinesepolandstarlatrialterminalairfieldchristiansffabishopverbgminasymptomspromotedrideplatformassignedgodsmidfieldercommandernounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrainairlinessubtopic 78subtopic 78subtopic 86subtopic 9subtopic 08subtopic 18subtopic 38cablepic 48cableteamracingboysdefendernavygenderlistingfflyingchanelsleaguev8tibetanncaateamssubtopic 08subtopic 19subtopic 38subtopic 48cableteamracingboysnbafootballsubtopic 09subtopic 19subtopic 39subtopic 39cableteamracingboysnbafootball	particle	starred						cathedral		
acid electronsproductions broadwayfounded launchedserbia hungarydecides leavescorridor bypassbattery bypassportugal speedsf1samesex increasedsubtopic 07 chinesesubtopic 17 polandsubtopic 27 subtopic 17subtopic 37 subtopic 37subtopic 47 lasubtopic 57 trialsubtopic 67 trialsubtopic 67 subtopic 67subtopic 77 subtopic 77subtopic 87 subtopic 87subtopic 97 subtopic 78verbsgmina ounssymptoms symptomspromoted charlesirrial arrestedterminal airlineairfield arrestedchristians squadronsf1 subtopic 78 subtopic 78subtopic 87 subtopic 79subtopic 70 subtopic 79subtopic 70 subtopic 79subtopic 70 subtopic 70subtopic 70 subtopic 70sub		season						papal		
electronsbroadwaylaunchedhungaryleavesbypassspeedsvenicehamiltonincreasedsubtopic 07subtopic 17subtopic 27subtopic 37lasubtopic 47subtopic 57subtopic 67subtopic 77subtopic 87subtopic 97chinesepolandstarlatrialterminalairfieldchristiansfifabishopverbgminasymptomspromotedrideplatformaassignedprayeruefacommanderoursevoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativmusclepuertotrianairlinemissionsgoodssubtopic 78subtopic 78subtopic 78subtopic 88subtopic 98subtopic 08subtopic 18subtopic 28subtopic 38subtopic 48subtopic 58subtopic 58subtopic 668subtopic 78subtopic 88subtopic 98seasonpeakedxwingcableteamracingboysnhafootballsuperlistingfflyingchannelsleaguev8tibetanncaancaateamssubtopic 09subtopic 19subtopic 39subtopic 49stadiumchassisguru<assistsbaltopic 39superlistingfflyingchannelsleaguev8tibetanncaancaateamssuperlistingfflying<	electron	cast					greater	santa		budget
subtopic 07 chinesesubtopic 17 polandsubtopic 27 starsubtopic 37 lasubtopic 47 trialsubtopic 57 terminalsubtopic 67 airfieldsubtopic 77 christianssubtopic 87 ffasubtopic 97 bishopverb verbsgminasymptoms symptomepromotedrideplatformairfield airfieldchristiansffa commandernounsvoivodeship bonebonesir sircriminal criminalnavy triancorpsdivineseriesubtopic 08 seasonsubtopic 18 peakedsubtopic 38 musclesubtopic 48 musclesubtopic 58 their triansubtopic 58 teamsubtopic 59 teamsubtopic 59 subtopic 59 subtopic 59 subtopic 59<	acid	productions	founded	serbia	decides	corridor	battery	portugal		samesex
chinesepolandstarIatrialterminalairfieldchristiansfifabishopverbgminasymptomspromotedrideplatformassignedprayeruefacommandeverbsdistrictsyndromecharlesarrestedairlinessquadronsgodsgodsmidfieldercommandenounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrainairlinesmissionsgospeldefendernavysubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gseasonpeakedxwingcableteamracingboysnba<	electrons	broadway		hungary			speeds	venice	hamilton	increased
chinesepolandstarIatrialterminalairfieldchristiansfifabishopverbgminasymptomspromotedrideplatformassignedprayeruefacommandeverbsdistrictsyndromecharlesarrestedairlinessquadronsgodsgodsmidfieldercommandenounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrainairlinesmissionsgospeldefendernavysubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gseasonpeakedxwingcableteamracingboysnba<	subtopic 07	subtopic 17	subtopic 27	subtopic 37	subtopic 47	subtopic 57	subtopic 67	subtopic 77	subtopic 87	subtopic 97
verbgminasymptomspromotedrideplatformassignedprayeruefacommandverbsdistrictsyndromecharlesarrestedairlinessyqudronsgodsmidfieldercommandernounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrainairlinemissionsgospeldefendernavysubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gsubtopic 1gsubtopic 2gsubtopic 2gsubtopi										
verbsdistrictsyndromecharlesarrestedairlinessquadronsgodsmidfieldercommandernounsvoivodeshipbonesircriminalnavycorpsdivineserieseminarygenderadministrativemusclepuertotrainairlineairlinemissionsgodsmidfieldernavysubtopic 0ssubtopic 1ssubtopic 2ssubtopic 3ssubtopic 4ssubtopic 5ssubtopic 5ssubtopic 6ssubtopic 7ssubtopic 7ssubtopic 7ssubtopic 9snbafootballsuperlistingfflyingchannelsleaguev8tibetanncaateamsnintendocontesttheoremenginesatelliteteamsccpupilstacklesathleticballgrammyequationsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gsubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gtacklestacklesathleticballgrammyequationsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gsubtopic 0gsubtopic 1gsubtopic 3gsubtopic 4gincomeequationministryfightatrestedsubtopic 1gsubtopic 3gsubtopic 4ghomerincomeequationministryfightatrestedsubtopic 3gs										
nouns gendervoivodeship administrativebonesircriminal puertonavy traincorpsdivineserieserieseminary navysubtopic 08 seasonsubtopic 18 peakedsubtopic 28 xsubtopic 38 wingsubtopic 64 cablesubtopic 68 teamsubtopic 68 racingsubtopic 78 	verbs				arrested				midfielder	
genderadministrativemusclepuertotrainairlinemissionsgospeldefendernavysubtopic 08subtopic 18subtopic 28subtopic 38subtopic 38subtopic 48subtopic 58subtopic 68subtopic 68subtopic 78subtopic 78subtopic 88subtopic 98seasonpeakedxwingcableleamracingboysnbafootballsuperlistingfflyingchannelsleaguev8tibetanncaateamsxmenrbpointsfighteraffiliatestadiumchassisguruassistsbasketballnintendocontesttheoremenginesatelliteteamsccpupilstacklesathleticballgrammyequationsubtopic 39subtopic 49subtopic 59subtopic 69subtopic 9subtopic 9subtopic 9subtopic 29subtopic 29subtopic 29subtopic 59subtopic 69subtopic 9subtopic 89subtopic 9universitymuseumspokenviolencehomerincomeequationministryfightarrestedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopal <td></td>										
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	gender	1								2
seasonpeakedxwingcableteamracingboysnbafootballsuperlistingfflyingchannelsleaguev8tibetanncaateamsxmenrbpointsfighteraffiliatestadiumchassisguruassistsbasketballnintendocontesttheoremenginesatelliteteamscccpupilstacklesathleticballgrammyequationsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9gsubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gincomeequationsubtopic 9gsubtopic 9gs										
superlistingfflyingchannelsleaguev8tibetanncaateamsxmenrbpointsfighteraffiliatestadiumchassisguruassistsbasketballnintendocontesttheoremenginesatelliteteamsccpupilstacklesathleticballgrammyequationsubmarinestoresplaysportsgirlsccachedclubssubtopic 09subtopic 19subtopic 29subtopic 39subtopic 49subtopic 59subtopic 69subtopic 79subtopic 89subtopic 99universitymuseumspokenviolencehomerincomeequationministryfightarrestedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
xmenrbpointsfighteraffiliatestadiumchassisguruassistsbasketballnintendocontesttheoremenginesatelliteteamsccpupilstacklesathleticballgrammyequationsubmainestoresplaysportsgirlscoachedclubssubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9guniversitymuseumspokenviolencehomerincomeequationministryfightarrestedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
nintendocontesttheoremenginesatelliteteamsccpupilstacklesathleticballgrammyequationsubmarinestoresplaysportsgirlscoachedclubssubtopic 0gsubtopic 1gsubtopic 2gsubtopic 3gsubtopic 4gsubtopic 5gsubtopic 6gsubtopic 6gsubtopic 7gsubtopic 8gsubtopic 9guniversitymuseumspokenviolencehomerincomeequationministryfightarcstedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
ballgrammyequationsubmarinestoresplaysportsrightscoachedclubssubtopic 09subtopic 19subtopic 29subtopic 39subtopic 49subtopic 59subtopic 59subtopic 69subtopic 79subtopic 79subtopic 89subtopic 99universitymuseumspokenviolencehomerincomeequationministryfightarrestedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
subtopic 0.9 university subtopic 1.9 museum subtopic 2.9 spoken subtopic 3.9 violence subtopic 4.9 homer subtopic 5.9 income subtopic 6.9 equation subtopic 7.9 ministry subtopic 8.9 fight subtopic 9.9 arrested students fiction dialects trial author median p meeting boxing told education gallery speakers al simpsons households theorem seminary salkes alleged professor science geological lebanon century 65 values episcopal decision killed										
universitymuseumspokenviolencehomerincomeequationministryfightarrestedstudentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
studentsfictiondialectstrialauthormedianpmeetingboxingtoldeducationgalleryspeakersalsimpsonshouseholdstheoremseminarystakesallegedprofessorsciencegeologicallebanoncentury65valuesepiscopaldecisionkilled										
education gallery speakers al simpsons households theorem seminary stakes alleged professor science geological lebanon century 65 values episcopal decision killed										
professor science geological lebanon century 65 values episcopal decision killed										
critical editor earths syria hart size vector theological 64 refused										
enter catus syna batt size vector uteological 04 refused	critical	editor	earths	syria	bart	size	vector	theological	64	refused

Table 4: Topic Hierarchy with branching factor 10 for Wikipedia corpus using SPLIT-SINGLE

Opt 0Uppe 1Uppe 2C++Microsoft WordSupply ChainJavaTeachersSupply ChainPHPTeachingBusiness StrategyCSSStudentsPurchasingC#Microsoft ExcelWarehousingCEnglishContract NegotiationMySQLCurriculum DesignInventory ManagementNETCoursesProcurementHTMLEditingNegotiationxMLAcademiaNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementMatlabDiscourseContinuous ImprovementJava EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJava EESubtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationJataLaboratory SkillsWarehousingApplicationsDissectionSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarkatingFFGALibrary ResearchMerchandisingFPGALibrary ResearchMerchandisingSubtopic 04Subtopic 14	topic 0	topic 1	topic 2
JavaTeachersSupply Chain ManagementPHPTeachingBusiness StrategyCSSStudentsPurchasingC#Microsoft ExcelWarehousingCEnglishContract NegotiationMySQLCurriculum DesignInventory ManagementNETCoursesProcurementHTMLEditingNegotiationXMLAcademiaNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementMatlabDiscourseContinuous ImprovementMatlabGraduate Record ExaminationsProcess ImprovementMatabGreek LifeNew Business Developmentsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeNew Business StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12Subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionSalespringCollectionsSalesMatlabDiscectionSalesSubtopic 03Subtopic 13Subtopic 23Electronic BeatoricsVirtual ReferenceMarketing<			
PHPTeachingBusiness StrategyCSSStudentsPurchasingC#Microsoft ExcelWarehousingCEnglishContract NegotiationMySQLCurriculum DesignInventory Management.NETCoursesProcurementHTMLEditingNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSustopic 20MatlabDiscourseContinuous ImprovementMatlabDiscourseContinuous ImprovementMatlabGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTorncatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationApplicationsDissectionSalesVHDLAcademic LibrariesBrandElectroic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectroicsVirtual ReferenceMarketingFPGALibrary Research <t< td=""><td></td><td></td><td></td></t<>			
CSSStudentsPurchasingC#Microsoft ExcelWarehousingCEnglishContract NegotiationMySQLCurriculum DesignInventory Management.NETCoursesProcurementHTMLEditingNegotiationsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionSalesVHDLAcademic LibrariesBrandEnbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFFGALibrary ResearchMerchanisingSubtopic 04subtopic 14 </td <td></td> <td></td> <td></td>			
C#Microsoft ExcelWarehousingCEnglishContract NegotiationMySQLCurriculum DesignInventory ManagementNETCoursesProcurementHTMLEditingNegotiationXMLAcademiaNew Business Developmentsubtopic 00subtopic 10Subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementMatlabGraduate Record ExaminationsProcess ImprovementMatlabGraduate Record ExaminationsProcess ImprovementJava EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFFGALibrary ResearchMerchandisingSubtopic 04subtopic 14Subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQL			
CEnglishContract NegotiationMySQLCurriculum DesignInventory Management.NETCoursesProcurementHTMLEditingNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13Subtopic 23ElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 04subtopic 14Subtopic 24PHPSchool PsychologyProcurenentCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogistics			
MySQL NETCurriculum Design CoursesInventory Management ProcurementHTMLEditingNegotiationXMLAcademiaNew Business Developmentsubtopic 00 Researchsubtopic 10 English Compositionsubtopic 20 Lean ManufacturingAlgorithmsBritish Literature DataSix Sigma Continuous Improvement ManufacturingAnalysisWorld Literature Graduate Record ExaminationsContinuous Improvement Manufacturing Process Improvementsubtopic 01 Java EEsubtopic 11 Greek Lifesubtopic 21 New Business DevelopmentMibernate SpringCounselor Education Greek LifeNew Business Development Marketing Strategy StrategyJava ECommunity Colleges Greek LifeNew Business Development Marketing StrategyJava ECommunity DevelopmentStrategic PlanningSubtopic 02 Microsoft SQL ServerScience Communication LogisticsLogisticsC# Human Physiology ClientInvertebratesFreightApplications DissectionShipping TransportationSubtopic 23Electronics VHDL PGASubtopic 14 Library ResearchSubtopic 23Subtopic 04 FPGASubtopic 14 Library ResearchSubtopic 24 ProcurementMarketing FPGACognitive Neuroscience LogisticsMarketing MySQL Cognitive NeuroscienceJavascriptCommunication DisordersSupply Chain Management			
NETCoursesProcurementHTMLEditingNegotiationXMLAcademiaNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 1subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationApplicationsDissectionShippingDataLaboratory SkillsWarehousingApplicationsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 04Subtopic 14Subtopic 24PHPSchool PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management <td>_</td> <td></td> <td></td>	_		
HTMLEditingNegotiationXMLAcademiaNew Business Developmentsubtopic 00subtopic 10subtopic 20ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementMatlabGraduate Record ExaminationsProcess ImprovementSubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementFPGALibrary ResearchMerchandising<		e	
XMLAcademiaNew Business Developmentsubtopic 0_0 subtopic 1_0 subtopic 2_0 ResearchEnglish CompositionLean ManufacturingAlgorithmsBritish LiteratureSix SigmaMatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 0_1 subtopic 1_1 subtopic 2_1 Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 0_2 subtopic 1_2 subtopic 2_2 Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 0_3 subtopic 1_3 subtopic 2_3 ElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingFPGALibrary ResearchMerchandisingFPGALibrary ResearchMerchandisingSubtopic 0_4 subtopic 1_4 subtopic 2_4 PHPSchool PsychologyProcurementFPGALibrary ResearchMerchandisingSubtopic 0_4			
$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$			
Research AlgorithmsEnglish Composition British LiteratureLean Manufacturing Six SigmaMatlabDiscourseSix SigmaMatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12Subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 04Subtopic 14Subtopic 24PHPSchool PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
Algorithms MatlabBritish Literature DiscourseSix Sigma Continuous ImprovementMatlabDiscourse World LiteratureContinuous ImprovementAnalysisWorld Literature Graduate Record ExaminationsManufacturing Process ImprovementDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11 Subtopic 21Subtopic 21 Marketing StrategyJava EECommunity Colleges Greek LifeNew Business Development Marketing StrategySpringCounselor Education First Year Experience JavaSales Management Business StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12 Science Communication Laboratory SkillsSubtopic 22 Microsoft SQL ServerC#Human Physiology Data Laboratory SkillsTransportation Shipping Shipping Client Invertebratessubtopic 03subtopic 13 Subtopic 23Subtopic 23 Secial Collections SalesVHDL ElectronicsAcademic Libraries Virtual ReferenceRetail Marketing MerchandisingSubtopic 04subtopic 14 School PsychologySubtopic 24 Procurement ProcurementPHPSchool Psychology Procurement CogisticsProcurement MerchanisingSubtopic 04Subtopic 14 Subtopic 24 PhPSchool Psychology Purchasing Library ResearchMySQL Cognitive Neuroscience JavascriptCommunication DisordersSupply Chain Management			
MatlabDiscourseContinuous ImprovementAnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23ElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingMurchandisingLibrary ResearchMerchandisingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
AnalysisWorld LiteratureManufacturingDataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23ElectronicsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
DataGraduate Record ExaminationsProcess Improvementsubtopic 01subtopic 11subtopic 21Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03Subtopic 13subtopic 23Electronica EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
subtopic 0_1 subtopic 1_1 subtopic 2_1 Java EECommunity CollegesNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 0_2 subtopic 1_2 subtopic 2_2 Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 0_3 subtopic 1_3 subtopic 2_3 Electroical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingSubtopic 0_4 subtopic 1_4 subtopic 2_4 PHPSchool PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Analysis		
Java EECommunity Colleges Greek LifeNew Business DevelopmentHibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
HibernateGreek LifeMarketing StrategySpringCounselor EducationSales ManagementTomcatFirst Year ExperienceBusiness StrategyJavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	subtopic 0 ₁	subtopic 1 ₁	
Spring TomcatCounselor Education First Year ExperienceSales Management Business StrategyJavaPersonality DevelopmentSubtopic 22JavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Java EE		
Toncat JavaFirst Year Experience Personality DevelopmentBusiness Strategy Strategic Planningsubtopic 02 Microsoft SQL Server C#subtopic 12 Science Communicationsubtopic 22 LogisticsMicrosoft SQL Server C#Science Communication Human PhysiologyLogisticsDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical Engineering Embedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Hibernate	Greek Life	Marketing Strategy
JavaPersonality DevelopmentStrategic Planningsubtopic 02subtopic 12subtopic 22Microsoft SQL ServerScience CommunicationLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Spring	Counselor Education	
subtopic 0_2 subtopic 1_2 subtopic 2_2 Microsoft SQL Server C#Science Communication Human PhysiologyLogisticsDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 0_3 subtopic 1_3 subtopic 2_3 Electrical EngineeringElectronic ResourcesRetailElectronicsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 0_4 School PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Tomcat	First Year Experience	Business Strategy
Microsoft SQL Server C#Science Communication Human PhysiologyLogisticsC#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Java	Personality Development	Strategic Planning
C#Human PhysiologyTransportationDataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	subtopic 0 ₂	subtopic 1 ₂	subtopic 2 ₂
DataLaboratory SkillsWarehousingApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Microsoft SQL Server	Science Communication	Logistics
ApplicationsDissectionShippingClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04Subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	C#	Human Physiology	Transportation
ClientInvertebratesFreightsubtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Data	Laboratory Skills	Warehousing
subtopic 03subtopic 13subtopic 23Electrical EngineeringElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Applications	Dissection	Shipping
Electrical Engineering Embedded SystemsElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Client	Invertebrates	Freight
Electrical Engineering Embedded SystemsElectronic ResourcesRetailEmbedded SystemsSpecial CollectionsSalesVHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	subtopic 0 ₃	subtopic 1 ₃	subtopic 2 ₃
VHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Electrical Engineering	Electronic Resources	
VHDLAcademic LibrariesBrandElectronicsVirtual ReferenceMarketingFPGALibrary ResearchMerchandisingsubtopic 04subtopic 14subtopic 24PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Embedded Systems	Special Collections	Sales
FPGALibrary ResearchMerchandisingsubtopic 0_4 subtopic 1_4 subtopic 2_4 PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			Brand
FPGALibrary ResearchMerchandisingsubtopic 0_4 subtopic 1_4 subtopic 2_4 PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	Electronics	Virtual Reference	Marketing
subtopic 0_4 subtopic 1_4 subtopic 2_4 PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	FPGA	Library Research	
PHPSchool PsychologyProcurementCSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	subtopic 0 ₄		
CSSAbnormal PsychologyPurchasingMySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management			
MySQLCognitive NeuroscienceLogisticsJavascriptCommunication DisordersSupply Chain Management	CSS		Purchasing
Javascript Communication Disorders Supply Chain Management			
		Health Psychology	

Table 5: Subset of hierarchy on LinkedIn member profiles. A 2-level hierarchy with branching factor 15 was learned using SPLIT-SINGLE. Each "document" was a member profile, and the terms in each document were explicitly or implicitly specified skills identified by LinkedIn (see http://www.linkedin.com/skills/). There were tens of millions of documents and tens of thousands of terms, with a training time of approximately one day using hundreds of mappers and reducers.