Articles

Publicaton source	Title	Year
SIGIR	Estimating the Query Difficulty for Information Retrieval	2010
KAIS	A Enhancing supervised bug localization with metadata and stack-trace	2020
KAIS	On the Relationship between Bug Reports and Queries for Text Retrieval-based Bug Localization	2021
IST	A Deep-Learning-Based Bug Priority Prediction Using RNN-LSTM Neural Networks	2021
IST	ManQ: Many-objective optimization-based automatic query reduction for IR-based bug localization	2020

Estimating the Query Difficulty for Information Retrieval -SIGIR.2010

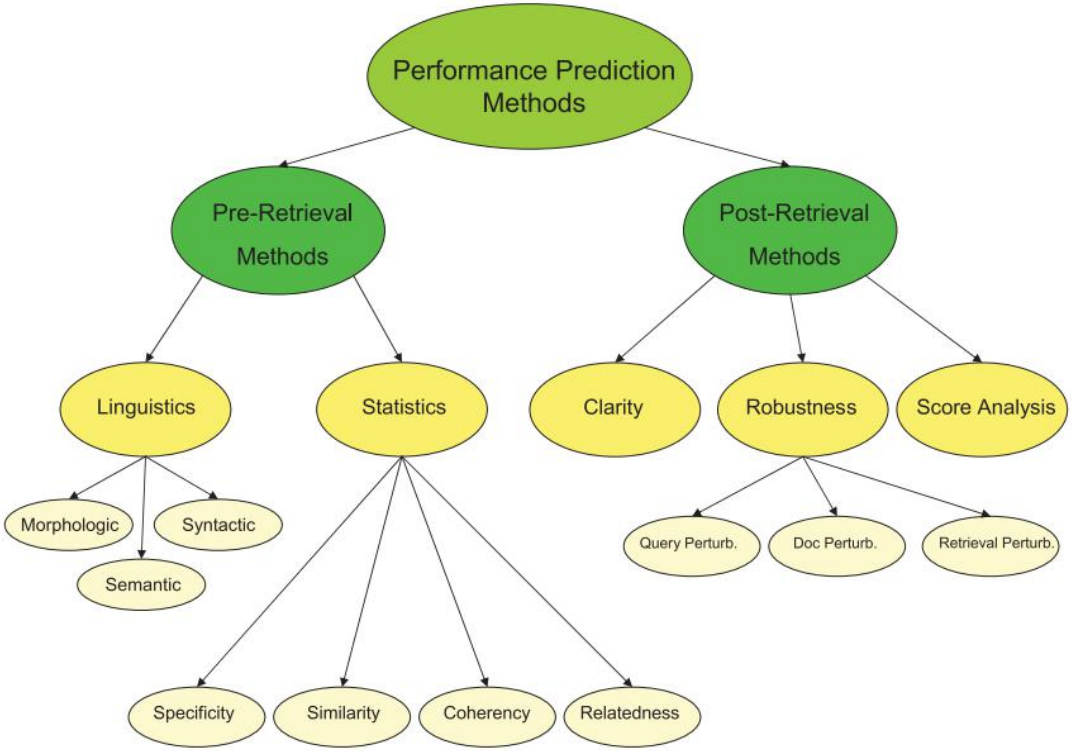


Figure 3.1: A general taxonomy of query performance prediction methods.

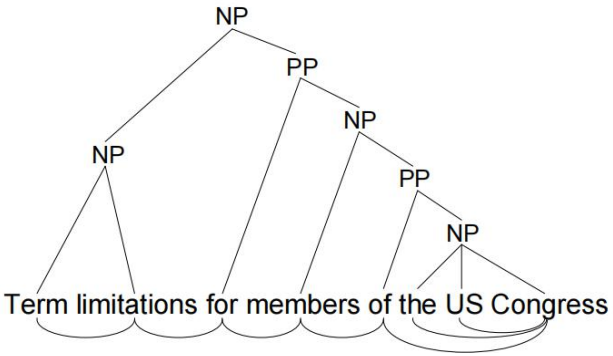
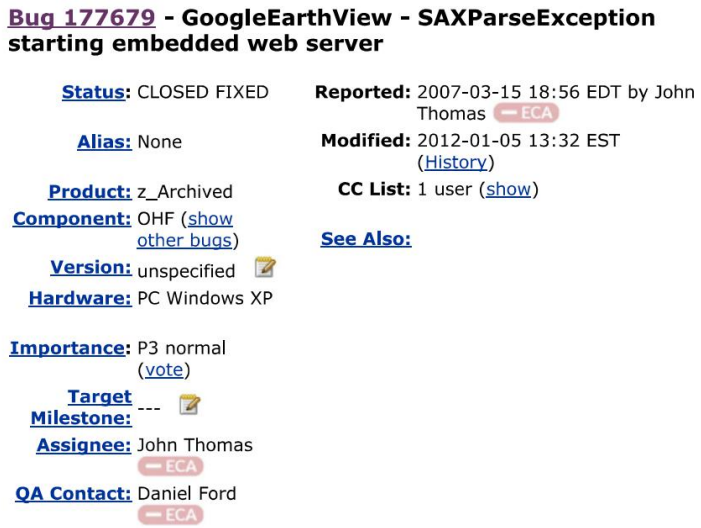


Figure 4.1: Syntactic parse tree

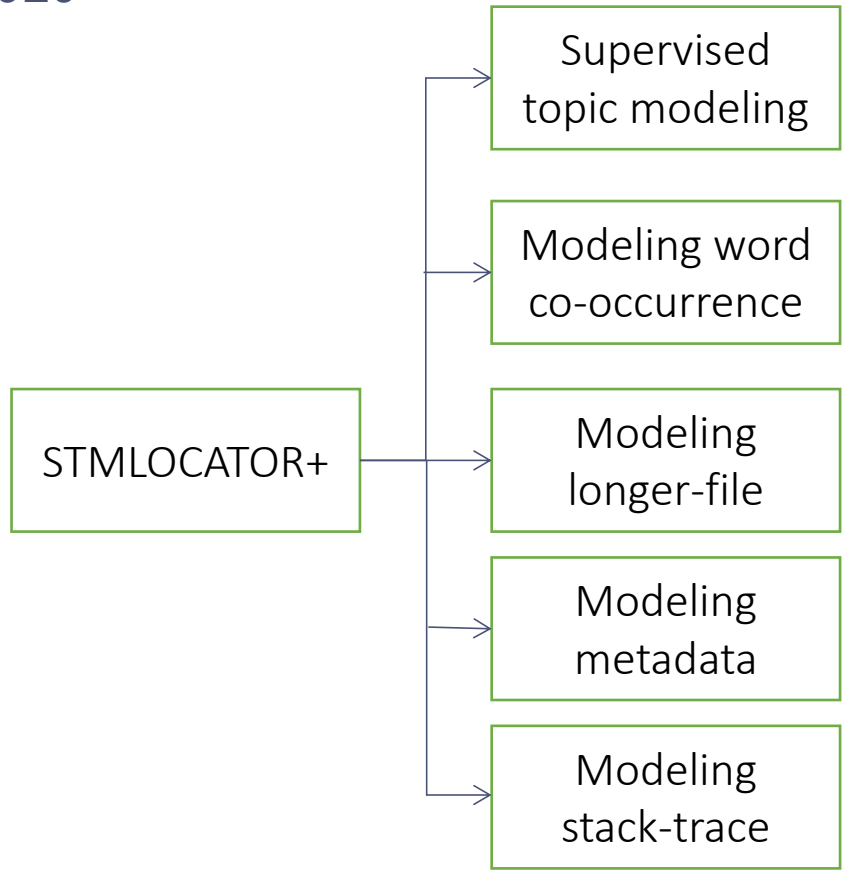
Enhancing supervised bug localization with metadata and stack-trace -KAIS.2020



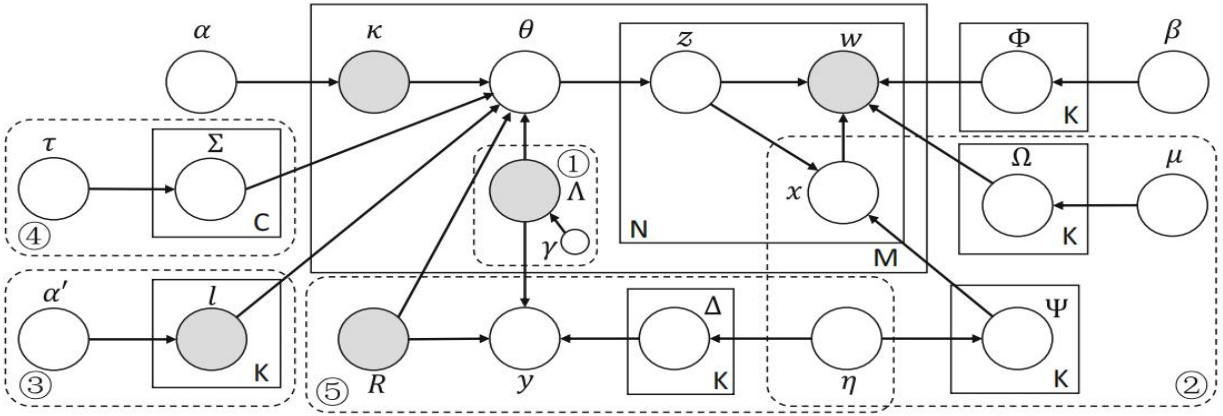
Bug 177679 - GoogleEarthView - SAXParseException starting embedded web server

Status: CLOSED FIXED	**Reported:** 2007-03-15 18:56 EDT by John Thomas ECA
Alias: None	**Modified:** 2012-01-05 13:32 EST (History)
Product: z_Archived	**CC List:** 1 user (show)
Component: OHF (show other bugs)	**See Also:**
Version: unspecified	
Hardware: PC Windows XP	
Importance: P3 normal (vote)	
Target Milestone: ---	
Assignee: John Thomas ECA	
QA Contact: Daniel Ford ECA	

Fig. 1 An example of metadata from a bug report



Enhancing supervised bug localization with metadata and stack-trace -KAIS.2020



M: # of bug reports
N: # of words
C: # of components
K: # topics/source files
w: observed word
z: topic of word w
θ: topic distribution
Λ: observed topics

l: observed file length
Φ: topic-word distribution
Ω: topic-word distribution
x: selection indicator
Ψ: selection distribution of x
y: selection indicator
κ: observed component
Δ: selection distribution of y

Σ: component-topic distribution
R: observed source files in stack-trace

① Supervised topic modeling
② Word co-occurrence phenomenon
③ Longer-file phenomenon
④ Meta-information constraints
⑤ Stack-trace tracking

Fig. 5 Graphical representation of STMLOCATOR+

On the Relationship between Bug Reports and Queries for Text Retrieval-based Bug Localization -KAIS.2021

Table 3 RQ_1: Comparison of queries Q_a and Q_{nH}. A pair of bold values indicates statistical significance between them at 95% confidence with at least a small effect size and * indicates medium effect sizes.

	Median Eff.		HITS@1		HITS@5		HITS@10		MAP		MRR	
Project	Q_a	Q_{nH}	Q_a	Q_{nH}	Q_a	Q_{nH}	Q_a	Q_{nH}	Q_a	Q_{nH}	Q_a	Q_{nH}
AspectJ	33	36	0.04	0.06	0.17	0.18	0.28	0.26	0.09	0.10	0.12	0.13
Birt	58	56	0.02	0.02	0.08	0.05	0.15	0.14	0.07	0.06	0.10	0.09
BookKeeper	2	3	0.29	0.21	0.79	0.67	0.83	0.75	0.38	0.33	0.51	0.41
Derby	**15**	**26**	0.16	0.12	0.39	0.31	0.43	0.39	0.17	0.16	0.25	0.21
JodaTime	**2***	**13***	0.14	0.14	**0.57**	**0.14**	**0.57**	**0.29**	0.38	0.21	0.38	0.21
Lucene	**2.5***	**11***	0.31	0.20	**0.69**	**0.33**	0.78	0.47	0.38	0.25	**0.47**	**0.28**
Mahout	**5***	**25***	**0.36**	**0.12**	**0.64**	**0.32**	**0.80**	**0.40**	**0.46***	**0.18***	**0.46***	**0.19***
OpenJpa	**5.5**	**18.5**	0.19	0.13	**0.50***	**0.25***	0.56	0.31	**0.32**	**0.15**	0.35	0.22
Pig	9	10.5	0.25	0.19	0.44	0.39	0.53	0.50	0.30	0.28	0.35	0.30
Solr	**2**	**7**	**0.33**	**0.18**	0.67	0.42	0.76	0.56	0.42	0.26	0.49	0.30
SWT	**3**	**5**	0.35	0.31	0.61	0.53	0.78	0.69	0.43	0.37	0.48	0.43
Tika	**2***	**16***	**0.43***	**0.10***	0.62	0.29	0.81	0.38	**0.38***	**0.16***	**0.53***	**0.21***
Tomcat	2	12.5	0.40	0.23	0.67	0.37	0.74	0.41	0.55	0.31	0.59	0.34
ZooKeeper	**2**	**5**	**0.41***	**0.17***	0.73	0.52	0.79	0.64	**0.49**	**0.28**	**0.55**	**0.32**
ZXing	2	6	0.36	0.31	0.57	0.46	0.64	0.54	0.45	0.36	0.49	0.40
Total	**9.6**	**16.7**	0.27	0.17	0.54	0.36	0.63	0.45	0.35	0.23	**0.41**	**0.27**

On the Relationship between Bug Reports and Queries for Text Retrieval-based Bug Localization -KAIS.2021

Table 4 $RQ_{2.1}$: Performance of *near-optimal queries* QGA_a and QGA_{nH} (using Lucene 2.9.4)

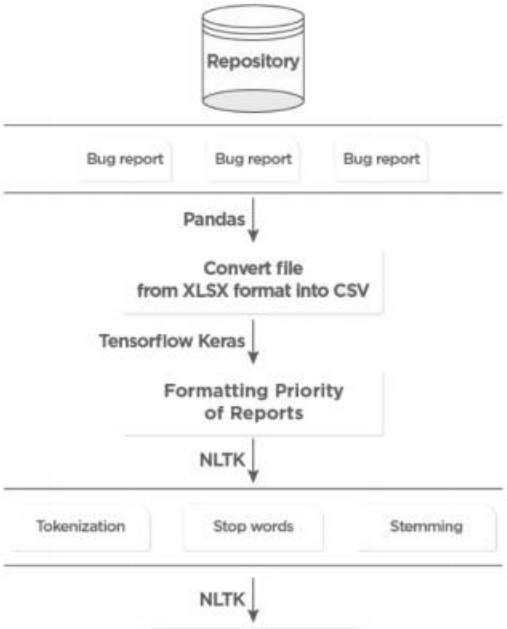
	Median Eff.		HITS@1		HITS@5		HITS@10		MAP		MRR	
Project	QGA_a	QGA_{nH}	QGA_a	QGA_{nH}	QGA_a	QGA_{nH}	QGA_a	QGA_{nH}	QGA_a	QGA_{nH}	QGA_a	QGA_{nH}
AspectJ	1	1	0.46	0.60	0.72	0.79	0.78	0.86	0.41	0.49	0.58	0.68
BookKeeper	1	1	0.88	0.75	1.00	0.96	1.00	0.96	0.56	0.48	0.92	0.85
Birt	1	1	0.34	0.35	0.50	0.51	0.55	0.57	0.43	0.43	0.61	0.64
Derby	1	1	0.55	0.53	0.67	0.65	0.67	0.73	0.40	0.40	0.61	0.60
JodaTime	1	1	0.57	0.57	1.00	0.86	1.00	0.86	0.79	0.73	0.79	0.73
Lucene	1	1	0.80	0.63	0.93	0.80	0.97	0.80	0.63	0.49	0.85	0.72
Mahout	1	1	0.88	0.56	0.96	0.72	1.00	0.76	0.84	0.52	0.91	0.63
OpenJpa	1	1	0.75	0.75	0.88	0.94	0.94	1.00	0.63	0.61	0.82	0.83
Pig	1	1	0.69	0.64	0.86	0.81	0.89	0.86	0.65	0.61	0.77	0.71
Solr	1	1	0.84	0.71	0.91	0.84	0.93	0.89	0.68	0.57	0.88	0.77
SWT	1	1	0.79	0.76	0.89	0.91	0.94	0.93	0.69	0.67	0.85	0.83
Tika	1	1	0.90	0.71	0.90	0.86	0.90	0.86	0.59	0.42	0.91	0.78
Tomcat	1	1	0.77	0.49	0.89	0.68	0.91	0.75	0.85	0.59	0.95	0.70
ZooKeeper	1	1	0.81	0.74	0.95	0.88	0.96	0.94	0.74	0.70	0.86	0.80
ZXing	1	1	0.93	0.85	1.00	0.92	1.00	1.00	0.91	0.86	0.95	0.90
Total	1	1	0.73	0.64	0.81	0.81	0.9	0.85	0.57	0.57	0.82	0.74

On the Relationship between Bug Reports and Queries for Text Retrieval-based Bug Localization -KAIS.2021

Table 15 Composition of original and near-optimal queries with respect to OB, EB, and S2R without localization hints by system

	Unoptimized				Near-Optimal			
system	OB	EB	S2R	Eff	OB	EB	S2R	Eff
AspectJ	52.58	12.82	14.11	92	51.53	12.32	14.49	12
Birt	32.51	13.56	45.36	599	37.13	15.31	47.56	93
Bookkeeper	43.56	30.09	0.00	13	43.64	32.85	0.00	2
Derby	41.47	25.24	23.08	169	42.83	29.32	22.42	31
JodaTime	34.92	12.56	8.53	30	35.81	11.74	9.57	3
Lucene	56.54	30.09	2.85	73	58.81	25.51	2.19	15
Mahout	51.92	25.24	0.00	234	58.85	17.61	0.00	15
OpenJpa	65.59	12.56	7.74	39	73.36	15.41	9.23	2
Pig	68.83	13.58	11.73	110	77.77	16.10	9.76	49
Solr	54.04	14.48	9.38	173	57.39	15.95	9.52	17
SWT	59.06	19.32	28.23	28	58.50	22.50	27.45	5
Tika	26.28	58.68	0.00	47	29.73	53.09	0.00	7
Tomcat	50.19	16.21	13.55	94	55.19	16.82	13.47	18
ZooKeeper	54.90	23.30	10.67	22	58.87	24.51	10.89	4
ZXing	16.43	8.07	24.44	20	18.22	9.26	25.53	2

A Deep-Learning-Based Bug Priority Prediction Using RNN-LSTM Neural Networks -IST.2021 -



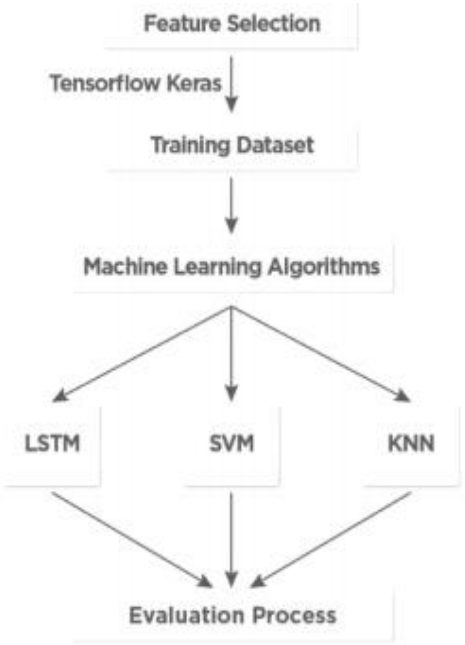


Figure 2: Proposed framework

A Deep-Learning-Based Bug Priority Prediction Using RNN-LSTM Neural Networks -IST.2021

Table 5: The top-30 keywords based on their frequency (sorted using NLTK)

Rank	Keyword	Rank	Keyword	Rank	Keyword
1	IOS	11	search	21	back
2	Android	12	seller	22	login
3	Screen	13	click	23	logo
4	app	14	App	24	account
5	incorrect	15	Api	25	network
6	message	16	backend	26	payment
7	product	17	chat	27	google
8	user	18	button	28	service
9	order	19	mobile	29	server
10	error	20	design	30	web

Table 6: Keywords classified based on the priority level

Keywords	Count of frequency	Priority level	Keywords	Count of frequency	Priority level
crash	186	high	color	29	low
error	159	high	inconsistent	22	low
icon	110	low	layout	21	low
photo	108	low	avatar	20	low
tab	65	low	placeholder	20	low
image	101	low	doesn't Work	16	high
menu	53	low	ux	16	low
design	52	low	toolbar	15	low
logo	47	low	textview	9	low
label	45	low	hot fix	9	high
title	34	low	failure	8	high

ManQ: Many-objective optimization-based automatic query reduction for IR-based bug localization -IST.2020

Table 1
The 15 objective functions of ManQ.

Component	Feature	Objective function
Maintaining Pre-retrieval QQPs	Coherency	$f_1(Q) = SumVAR(Q) = \sum_{q \in Q} \sqrt{\frac{\sum_{q \in D_t} (w(t,d) - \overline{w_t})^2}{df(t)}}$
	Similarity	$f_2(Q) = SumSCQ(Q) = \sum_{q \in Q} (1 + log(tf(q, D))) \cdot idf(q)$
	Term Relatedness	$f_3(Q) = SumPMI(Q) = \sum_{q_1, q_2 \in Q} log \frac{p_{q_1,q_2}(D)}{p_{q_1}(D) \cdot p_{q_2}(D)}$
	Specificity	$f_4(Q) = SumIDF(Q) = \sum_{q \in Q} log \frac{\lvert D \rvert}{\lvert D_q \rvert}$
		$f_5(Q) = SumICTF(Q) = \sum_{q \in Q} log \frac{\lvert D \rvert}{tf(t,D)}$
		$f_6(Q) = SumENT(Q) = \sum_{q \in Q} \sum_{d \in D_q} \frac{tf(t,d)}{tf(t,D)} \cdot log_{\lvert D \rvert} \frac{tf(t,d)}{tf(t,D)}$
		$f_7(Q) = 1 - QS(Q) = 1 - \frac{\lvert \cup_{q \in Q} D_q \rvert}{\lvert D \rvert}$
		$f_8(Q) = SCS(Q) = \sum_{q \in Q} D_q(Q) \cdot \log \frac{P_q(Q)}{P_q(D)}$
Maintaining Important Keywords	Obs. & Exp. Behavior*	$f_9(Q) = \frac{1}{\lvert OEB \rvert} \cdot \sum_{q \in Q} inOEB(q)$
	File Name	$f_{10}(Q) = HasFileName(Q)$
	Keywords in Summary	$f_{11}(Q) = \frac{1}{4} \cdot \sum_{q \in Q} isKeyword(q)$
	Informative Part-of-Speech	$f_{12}(Q) = \frac{1}{\lvert Q \rvert} \cdot \sum_{q \in Q} posScore(q) \begin{cases} 1.0, & q \ is \ Noun \\ 0.8, & q \ is \ Verb \ or \ Adjective \\ 0.4, & q \ is \ Adverb \end{cases}$
Maintaining Initial Information	Context of Sentence (S)	$f_{13}(Q) = SumQueryPMI(Q) = \sum_{q_1, q_2 \in Q} log \frac{p_{t_1,t_2}(S)}{p_{t_1}(S) \cdot p_{t_2}(S)}$
	Similarity with Initial Query	$f_{14}(Q) = CosineSimilarity(Q, O)$
Minimizing Query Length	Ratio of Query Length	$f_{15}(Q) = 1 - \frac{\lvert Q \rvert}{\lvert O \rvert}$

谢谢观看