



PROBLEM

MATERIAL Home About Login

189 results for "book": (biblia kazi kitabu)

Score	id	Links	English Excerpts	Swahili Excerpts
1	8358564	en, sw	<ul style="list-style-type: none"> tmj hospital and a team of doctors have done a great job for the sake of this project for preparing the swahili kidney book and a website for kidney patients. he said that a swahili book and a website about kidney infection will benefit millions of kidney patients and approximately 100 million swahili users in fighting kidney infection kigwanga said that the book aims at providing basic information to patients and their families about all the kidney problems the acting director of tmj, doctor tayabai jafferje said the launch of the website will enable tanzanians to get various information about the infection 	<ul style="list-style-type: none"> aliciinda kitabu hicho, namba waziri wa afya, mamedeleo ya jamii, jinsia, wazee na watoto, dk hami kigwagwala, alisema, "upakaraji wa kitabu cha figo na tovuti kwa kiswahili, kinachongwa na wata wengi ni hasia kutwa katika kuzua na kutibu magorwa ya figo hospital ya tmj na timu ya madaaktari wamelanya kitabu kubwa kwa ajili ya mtradi huu kwa kuweza kuandaa kitabu cha figo na tovuti kwa wogorwa wa figo kwa kiswahili, alisema kitabu na tovuti kuhusu ugonjwa wa figo kwa kiswahili vitawafuata mamalioni ya wogorwa wa figo na takribani watumaji wa kiswahili milioni 100 kupambana na ugonjwa wa figo kigwanga alisema kwamba kitabu hicho kinatenga kutoa habari za mzungu kwa wogorwa na familia zao kuhusu mizozo yote ya figo
0.59	44727963	en, sw	<ul style="list-style-type: none"> but we see that after lowering the price the citizens are not benefiting because of that they are charged very highly because the contractors themselves are using to go to the villages to do the jobs after finding it explained to yusuf to give me like a week i will get in that book number one in the book of contract need you to use the pillar of thirteen frameworks 	<ul style="list-style-type: none"> lakini tunaona kuwa baada ya kushusha bei wanaochi hawafadiki kwa sababu ya hivyo wanapigwa bei kubwa sana kwa sababu wale mcontractors wenyewe kamili wanatumia kuenda vijiji kufanya zile kitabu baada ya kukipata rikamweza vya ul anpe kama wiki mimi nitangia kwenye hicho kitabu namba moja kwenye kitabu cha contract tunataka utumie nguzo za mii kumi na tatu

- Search English queries
- Documents are indexed in other languages
- The most common approaches are query translation and document translation
- query translation by **Dictionary** and document translation by **Machine Translation**

QT AND DICTIONARY

' $q_1 q_2$ ', q_3

Fig. 1 The proposed phrase-based query translation. t_{ij} is the j -th translation for q_i . t_{12} and t_{22} are phrases in target language.

- build all combinations and take the **maximum**
- $\# \max([\#1(t_{11}, t_{22}), t_{32}], [\#1(t_{12}, t_{22}), t_{31}])$

DT AND PROXIMITY

' $q_1 q_2$ ', q_3

Fig. 2 The proposed query hit proximity approach. A query hit should have q_1 and q_2 adjacent to each other. Based on how far are the query terms are located in each document $f(q, d)$ gets lower score. This metric keeps the relative order of the query terms for matching.

- Find different **combinations** of query terms' positions
- Find query hits that keep the **order** of the query
- take the maximum possible hits
- $\# \max([1, 2, 23], [1, 2, 29], [15, 35, 71])$
- we can also have partial matching where a part of the query appears in the document
- then we can combine $p(q|d)$ based on language modeling:

$$s(d) = \alpha * p(q|d) + (1 - \alpha) * f(q, d) \quad (1)$$

$$f(q, d) = -\log\left(\sum_{i=0}^{|q|} \|q_i, q_{i+1}\|_d\right) \quad (2)$$

EXPERIMENTAL SETUP

- We used Indri retrieval system
- Europarl for CLEF collections and Bitext for surprise languages as training data for MT
- We used hard cutoff of 20 for surprise languages and then ran a tuning from 1 to 20 to find a suitable cutoff for the collection
- Wiktionary** is used as a bilingual dictionary
- Neural Machine Translation** is used for document translation
- Indri scores, word embedding, IDF, document length, and query length are features for **soft cutoff**
- We used normalization and lower casing for all collections

Table 1. Collection Statistics

ID	Lang.	Collection	Queries	#docs	#qrels
SP	Spanish	EFE 1994	CLEF 2002, topics 91-140	215,738	2,854
de	German	Frankfurter Rundschau 94, SDA 94, Der Spiegel 94-95	CLEF 2002-03, topics 91-140	225,371	1,938
fr	French	Le Monde 94, SDA French 94-95	CLEF 2002-03, topics 251-350	129,806	3,524
sw	Swahili	Analysis	300 constrained queries	471	390
tl	Tagalog	Analysis	300 constrained queries	462	233

EXPERIMENTS

- Document translation get's consistently better results than Query translation
- Phrase based retrieval get's as same result as synonym operator
- Cutoff is very important for AQWV but not MAP
- Wiktionary provides multiple choices for matching but MT provides only one option
- System combination** is a necessary step
- We should consider variance around the best cutoff for each collection
- Dictionary coverage is very important since a lot of queries have not translations

Table 2. Experimental Results

Lang	ANALYSIS/DEV	QT/DT	Phrase	Morph	Reranking	MAP	P@10	AQWV
SW	ANALYSIS	QT	no	no	no	0.213	0.0805	0.1025
SW	ANALYSIS	DT	no	no	no	0.3162	0.1085	0.1877
SW	ANALYSIS	QT	yes	no	no	0.212	0.0727	0.1313
SW	ANALYSIS	DT	yes	no	no	0.3303	0.107	0.2606
TL	ANALYSIS	QT	no	no	no	0.2469	0.0656	0.202
TL	ANALYSIS	DT	no	no	no	0.5285	0.1286	0.3389
TL	ANALYSIS	QT	yes	no	no	0.24	0.0641	0.2324
TL	ANALYSIS	DT	yes	no	no	0.5621	0.1296	0.4282