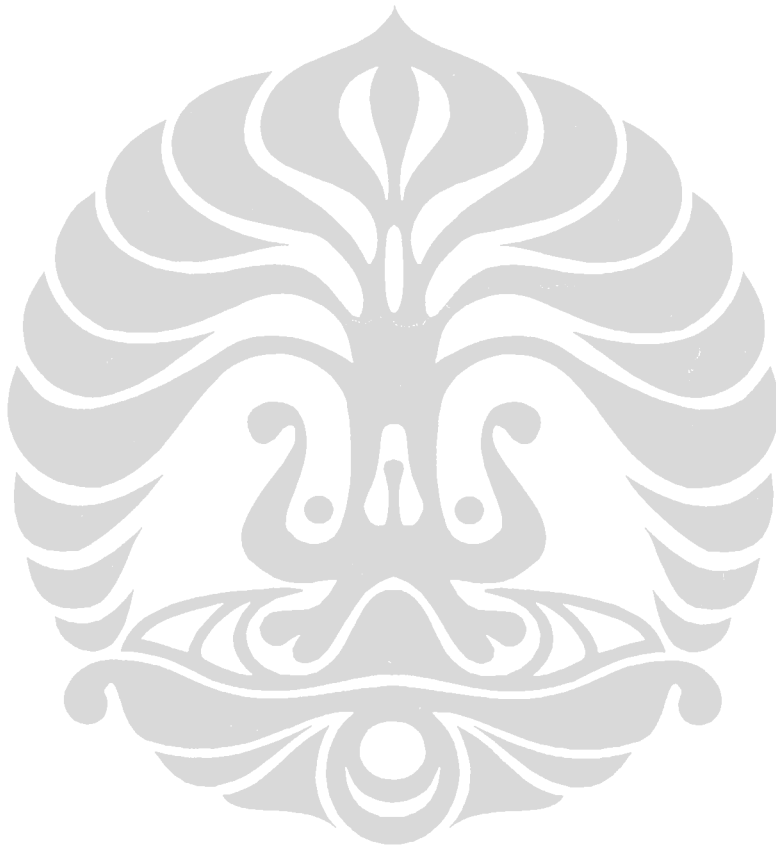


Appendix A

Sample Resources



C.1 Sample of Parallel Corpus 1

A.1.1 Sample of Article Pairs

- a. Good article translations

```

<DOC>
<DOCID>ANT-ID-00048</DOCID>
<SO>Antara</SO>
<DATE>20 September 2006 13:27</DATE>
<TITLE>Amien : Keputusan Eksekusi Tibo Cs Harus Dihormati</TITLE>
<TEXT>
Bandung (ANTARA News) - Mantan Ketua MPR, Amien Rais, mengatakan keputusan eksekusi terhadap terpidana mati kerusuhan Poso, Fabianus Tibo Cs, pada Kamis (21/9) atau Jumat (22/9) dinihari, merupakan keputusan yang harus dihormati.

"Saya kira kepolisian atau kejaksaan, cukup matang dan berpikir panjang dalam mengambil keputusan final soal eksekusi itu", katanya, se usai menjadi pembicara dalam acara Serikat Pekerja Pos Indonesia, di Bandung, Rabu.

Menurut dia, memang ada pihak-pihak dari luar yang mencoba untuk mengintervensi eksekusi terhadap Fabianus Tibo Cs, tapi yang paling penting kedaulatan hukum Indonesia harus dipertahankan.

Ia mengatakan lembaga hukum Indonesia telah berpikir panjang dan mengambil konklusi yang final bagi siapa saja yang telah membunuh di negeri tidak terlepas dari agamanya, sukunya, rasnya, atau latar belakang kedaerahannya, sudah pantas mendapatkan hukuman yang setimpal.

"Masalah Tibo ini hal yang sensitif, tapi negara kita adalah negara hukum, maka kita hormati keputusan hukum itu", katanya.

Meski demikian, ia mengingatkan dalam pemberian keputusan hukum, jangan sampai terjadi diskriminasi, namun intinya siapa yang sudah melakukan pembunuhan, maka pantas untuk diberikan hukuman yang setimpal.

Sedangkan perihal nama-nama yang diberikan oleh Fabianus Tibo yang menjadi pelaku dalam kasus kerusuhan Poso, menurut Amien pihak kepolisian harus mengungkap nama-nama yang diberikan Tibo Cs.

Amien Rais juga menyoroti persoalan keputusan hukum di tanah air yang bisa dinetralisir atau dibuyarkan, seperti kadang-kadang seorang kriminal yang sudah diputuskan hukuman penjara selama sepuluh tahun, tapi tiga tahun sudah ada di lapangan golf. (*)
</TEXT>
</DOC>

```

Figure C.1.1 Indonesian Article for Good Translation Sample

```

<DOC>
<DOCID>ANT-EN-00048</DOCID>
<SO>Antara</SO>
<DATE>20 September 2006 21:51</DATE>
<TITLE>Decision to execute Tibo should be respected, says Amien Rais</TITLE>
<TEXT>
Bandung, West Java (ANTARA News) - Former chairman of the People`s Consultative
Assembly (MPR) Amien Rais said the judicial authorities` decision to execute Fabianus
Tibo et al on Thursday (Sept 21) for involvement in the May 2000 incident in Poso,
Central Sulawesi, was something that should be respected.

"I think the police and prosecutors have taken the the final decision based on
sufficiently thorough considerations," Amien said here Wednesday.

According to Amien, there were persons trying to intervere in the execution of Tibo and
his friends but the most important thing was that the sovereignty of Indonesia`s laws
must be maintained. He said the judicial authorities had prudently considered and
made a final conclusion on giving proportional punishment to any killer in the country
regardless of his or her religion, group, race and ethnicity.

"The Tibo issue is sensitive. However, our country is a country adhering to the rule of
law so we have to respect the legal decision," he said.

On the the names of persons Fabianus Tibo had claimed were really responsible for the
Poso incident, Amien said the police must reveal them. (*)
</TEXT>
</DOC>

```

Figure C.1.2 English Article for Good Translation Sample

b. Poor article translations

```

<DOC>
<DOCID>ANT-ID-00026</DOCID>
<SO>Antara</SO>
<DATE>9 September 2006 19:54</DATE>
<TITLE>Bom Meledak di Dekat Markas TNI Poso</TITLE>
<TEXT>
Palu (ANTARA News) - Sebuah bom meledak di dekat markas bantuan Batalion 714
Poso di jalan Tabatoki Kelurahan Kawua, kota Poso, Sulawesi Tengah (Sulteng), Sabtu
malam.

Ledakan yang terjadi sekitar pukul 20:25 Wita itu melukai seorang warga bernama
Nela (20), warga Kawua.

Lokasi ledakan hanya berjarak sekitar 100 meter dari markas bantuan Batalion 714
Poso.

```

```
(*)
</TEXT>
</DOC>
```

Figure C.1.3 Indonesian Article for Poor Translation Sample

```
<DOC>
<DOCID>ANT-EN-00026</DOCID>
<SO>Antara</SO>
<DATE>17 Oktober 2006 0:26</DATE>
<TITLE>Another bomb blast rocks Poso</TITLE>
<TEXT>
Jakarta (ANTARA News) - A bomb blast again disturbed the composure of residents in
the former sectarian conflict-torn Poso city, Central Sulawesi, on Monday evening.
The blast, this time happened at 19:15 local time in an empty house in Lorong Dolidi
Dawa, Pulau Nias Street, Poso district, abandoned by its owner after being gutted by fire
in the past.
"There are no casualties and material losses in the blast," Wirabuna Regional Military
Commander May. Gen. Arief Budi Sampurno told ANTARA News here Monday evening.
Until now the police is still trying to find the cause of the blast, he added.
In the period following the execution of Fabianus Tibo, Dominggusda Silva and Marinus
Riwu, there have been six bomb blasts. At least one resident was injured in the bomb
blasts.
On Monday morning, Reverend Irianto Kongkoli was found dead, and the police were
still after the killer.
Commander Arief further explained that there was no extra security precautions in the
post murder period.
The Indonesian National Defence Forces (TNI) was ready to take extra security
precautions and even to help the police in unravelling the latest case of the murder of
reverend Kongkoli.
The National Police Chief said the murder may be linked to the execution of Tibo and
his friends.(*).
</TEXT>
</DOC>
```

Figure C.1.4 English Article for Poor Translation Sample

A.1.2 List of Random Set Terms

Table A.1 List of Random Set Terms of P₁₀₀

No	Term	No	Term	No	Term	No	Term
1	woman	26	limit	51	chen	76	commander
2	integral	27	return	52	welcome	77	booked
3	before	28	bi	53	faster	78	especially
4	verified	29	19	54	toxic	79	presence
5	empowerment	30	measure	55	committee	80	low
6	ask	31	financial	56	inflation	81	promoter
7	double	32	western	57	first	82	fighting
8	valid	33	governor	58	cultural	83	join
9	imperial	34	mind	59	expressed	84	industrial
10	responsibility	35	quite	60	funded	85	boat
11	weather	36	rupiah	61	london	86	urgent
12	impossible	37	put	62	result	87	reduction
13	heat	38	reach	63	being	88	sixth
14	globe	39	saw	64	tournament	89	afternoon
15	light	40	sporadic	65	seed	90	side
16	registered	41	china	66	proved	91	election
17	african	42	opponent	67	number	92	technical
18	analysis	43	mozambique	68	rest	93	ranked
19	overseas	44	necessary	69	rebound	94	hitting
20	culture	45	willing	70	controversial	95	expected
21	ensure	46	try	71	signing	96	governing
22	scene	47	nearly	72	crossing	97	running
23	television	48	absolutely	73	album	98	beating
24	destroyed	49	moslem	74	severe	99	away
25	someone	50	league	75	labor	100	accused

A.1.3 List of Top Score Set Terms

Table A.2 List of Top Score Set Terms of P₁₀₀

No	Term	No	Term	No	Term	No	Term
1	said	26	billion	51	him	76	pounds
2	news	27	should	52	shares	77	very
3	million	28	security	53	rupiah	78	led
4	indonesia	29	while	54	since	79	well
5	indonesian	30	reported	55	foreign	80	trade

6	year	31	new	56	friday	81	now
7	government	32	production	57	island	82	australian
8	jakarta	33	percent	58	told	83	peace
9	two	34	tuesday	59	director	84	monday
10	world	35	israel	60	former	85	many
11	per	36	time	61	sunday	86	like
12	bank	37	day	62	quarter	87	java
13	president	38	workers	63	air	88	asked
14	people	39	our	64	timor	89	years
15	here	40	state	65	match	90	third
16	oil	41	military	66	make	91	four
17	first	42	economic	67	australia	92	group
18	during	43	second	68	number	93	compared
19	gas	44	wednesday	69	month	94	asia
20	minister	45	meeting	70	sent	95	cooperation
21	country	46	including	71	win	96	film
22	company	47	national	72	thursday	97	international
23	countries	48	added	73	members	98	police
24	down	49	around	74	end	99	price
25	east	50	before	75	against	100	cup

C.2 Sample of Parallel Corpus 2

C.2.1 Sample of Bible Chapters and Verses

- a. Sample of an English Bible Chapter and its verses

Table A.3 English Bible: Book Genesis Chapter 1

Chapter	Verse	Passage
1	1	At the first God made the heaven and the earth.
1	2	And the earth was waste and without form; and it was dark on the face of the deep: and the Spirit of God was moving on the face of the waters.
1	3	And God said, Let there be light: and there was light.
1	4	And God, looking on the light, saw that it was good: and God made a division between the light and the dark,
1	5	Naming the light, Day, and the dark, Night. And there was evening and there was morning, the first day.
1	6	And God said, Let there be a solid arch stretching over the waters, parting the waters from the waters.
1	7	And God made the arch for a division between the waters which were under the arch and those which were over it: and it was so.
1	8	And God gave the arch the name of Heaven. And there was evening and there was morning, the second day.

Chapter	Verse	Passage
1	9	And God said, Let the waters under the heaven come together in one place, and let the dry land be seen: and it was so.
1	10	And God gave the dry land the name of Earth; and the waters together in their place were named Seas: and God saw that it was good.
1	11	And God said, Let grass come up on the earth, and plants producing seed, and fruit-trees giving fruit, in which is their seed, after their sort: and it was so.
1	12	And grass came up on the earth, and every plant producing seed of its sort, and every tree producing fruit, in which is its seed, of its sort: and God saw that it was good.
1	13	And there was evening and there was morning, the third day.
1	14	And God said, Let there be lights in the arch of heaven, for a division between the day and the night, and let them be for signs, and for marking the changes of the year, and for days and for years:
1	15	And let them be for lights in the arch of heaven to give light on the earth: and it was so.
1	16	And God made the two great lights: the greater light to be the ruler of the day, and the smaller light to be the ruler of the night: and he made the stars.
1	17	And God put them in the arch of heaven, to give light on the earth;
1	18	To have rule over the day and the night, and for a division between the light and the dark: and God saw that it was good.
1	19	And there was evening and there was morning, the fourth day.
1	20	And God said, Let the waters be full of living things, and let birds be in flight over the earth under the arch of heaven.
1	21	And God made great sea-beasts, and every sort of living and moving thing with which the waters were full, and every sort of winged bird: and God saw that it was good.
1	22	And God gave them his blessing, saying, Be fertile and have increase, making all the waters of the seas full, and let the birds be increased in the earth.
1	23	And there was evening and there was morning, the fifth day.
1	24	And God said, Let the earth give birth to all sorts of living things, cattle and all things moving on the earth, and beasts of the earth after their sort: and it was so.
1	25	And God made the beast of the earth after its sort, and the cattle after their sort, and everything moving on the face of the earth after its sort: and God saw that it was good.
1	26	And God said, Let us make man in our image, like us: and let him have rule over the fish of the sea and over the birds of the air and over the cattle and over all the earth and over every living thing which goes flat on the earth.
1	27	And God made man in his image, in the image of God he made him: male and female he made them.
1	28	And God gave them his blessing and said to them, Be fertile and have increase, and make the earth full and be masters of it; be rulers over the fish of the sea and over the birds of the air and over every living thing moving on the earth.
1	29	And God said, See, I have given you every plant producing seed, on the face of all the earth, and every tree which has fruit producing seed: they will be for your food:
1	30	And to every beast of the earth and to every bird of the air and every living thing moving on the face of the earth I have given every green plant for food: and it was so.

Chapter	Verse	Passage
1	31	And God saw everything which he had made and it was very good. And there was evening and there was morning, the sixth day.

b. Sample of the corresponding Indonesian Bible Chapter and its verses

Table A.4 Indonesian Bible: Book Kejadian Chapter 1

Chapter	Verse	Passage
1	1	Pada mulanya Allah menciptakan langit dan bumi.
1	2	Bumi belum berbentuk dan kosong; gelap gulita menutupi samudera raya, dan Roh Allah melayang-layang di atas permukaan air.
1	3	Berfirmanlah Allah: "Jadilah terang." Lalu terang itu jadi.
1	4	Allah melihat bahwa terang itu baik, lalu dipisahkan-Nyalah terang itu dari gelap.
1	5	Dan Allah menamai terang itu siang, dan gelap itu malam. Jadilah petang dan jadilah pagi, itulah hari pertama.
1	6	Berfirmanlah Allah: "Jadilah cakrawala di tengah segala air untuk memisahkan air dari air."
1	7	Maka Allah menjadikan cakrawala dan Ia memisahkan air yang ada di bawah cakrawala itu dari air yang ada di atasnya. Dan jadilah demikian.
1	8	Lalu Allah menamai cakrawala itu langit. Jadilah petang dan jadilah pagi, itulah hari kedua.
1	9	Berfirmanlah Allah: "Hendaklah segala air yang di bawah langit berkumpul pada satu tempat, sehingga kelihatan yang kering." Dan jadilah demikian.
1	10	Lalu Allah menamai yang kering itu darat, dan kumpulan air itu dinamainya laut. Allah melihat bahwa semuanya itu baik.
1	11	Berfirmanlah Allah: "Hendaklah tanah menumbuhkan tunas-tunas muda, tumbuh-tumbuhan yang berbiji, segala jenis pohon buah-buahan yang menghasilkan buah yang berbiji, supaya ada tumbuh-tumbuhan di bumi." Dan jadilah demikian.
1	12	Tanah itu menumbuhkan tunas-tunas muda, segala jenis tumbuh-tumbuhan yang berbiji dan segala jenis pohon-pohonan yang menghasilkan buah yang berbiji. Allah melihat bahwa semuanya itu baik.
1	13	Jadilah petang dan jadilah pagi, itulah hari ketiga.
1	14	Berfirmanlah Allah: "Jadilah benda-benda penerang pada cakrawala untuk memisahkan siang dari malam. Biarlah benda-benda penerang itu menjadi tanda yang menunjukkan masa-masa yang tetap dan hari-hari dan tahun-tahun,
1	15	dan sebagai penerang pada cakrawala biarlah benda-benda itu menerangi bumi." Dan jadilah demikian.
1	16	Maka Allah menjadikan kedua benda penerang yang besar itu, yakni yang lebih besar untuk menguasai siang dan yang lebih kecil untuk menguasai malam, dan menjadikan juga bintang-bintang.
1	17	Allah menaruh semuanya itu di cakrawala untuk menerangi bumi,
1	18	dan untuk menguasai siang dan malam, dan untuk memisahkan terang dari gelap. Allah melihat bahwa semuanya itu baik.
1	19	Jadilah petang dan jadilah pagi, itulah hari keempat.

Chapter	Verse	Passage
1	20	Berfirmanlah Allah: "Hendaklah dalam air berkeriapan makhluk yang hidup, dan hendaklah burung beterbangan di atas bumi melintasi cakrawala."
1	21	Maka Allah menciptakan binatang-binatang laut yang besar dan segala jenis makhluk hidup yang bergerak, yang berkeriapan dalam air, dan segala jenis burung yang bersayap. Allah melihat bahwa semuanya itu baik.
1	22	Lalu Allah memberkati semuanya itu, firman-Nya: "Berkembangbiaklah dan bertambah banyaklah serta penuhilah air dalam laut, dan hendaklah burung-burung di bumi bertambah banyak."
1	23	Jadilah petang dan jadilah pagi, itulah hari kelima.
1	24	Berfirmanlah Allah: "Hendaklah bumi mengeluarkan segala jenis makhluk yang hidup, ternak dan binatang melata dan segala jenis binatang liar." Dan jadilah demikian.
1	25	Allah menjadikan segala jenis binatang liar dan segala jenis ternak dan segala jenis binatang melata di muka bumi. Allah melihat bahwa semuanya itu baik.
1	26	Berfirmanlah Allah: "Baiklah Kita menjadikan manusia menurut gambar dan rupa Kita, supaya mereka berkuasa atas ikan-ikan di laut dan burung-burung di udara dan atas ternak dan atas seluruh bumi dan atas segala binatang melata yang merayap di bumi."
1	27	Maka Allah menciptakan manusia itu menurut gambar-Nya, menurut gambar Allah diciptakan-Nya dia; laki-laki dan perempuan diciptakan-Nya mereka.
1	28	Allah memberkati mereka, lalu Allah berfirman kepada mereka: "Beranakcuculah dan bertambah banyak; penuhilah bumi dan taklukkanlah itu, berkuasalah atas ikan-ikan di laut dan burung-burung di udara dan atas segala binatang yang merayap di bumi."
1	29	Berfirmanlah Allah: "Lihatlah, Aku memberikan kepadamu segala tumbuh-tumbuhan yang berbiji di seluruh bumi dan segala pohon-pohonan yang buahnya berbiji; itulah akan menjadi makananmu.
1	30	Tetapi kepada segala binatang di bumi dan segala burung di udara dan segala yang merayap di bumi, yang bernyawa, Kuberikan segala tumbuh-tumbuhan hijau menjadi makanannya." Dan jadilah demikian.
1	31	Maka Allah melihat segala yang dijadikan-Nya itu, sungguh amat baik. Jadilah petang dan jadilah pagi, itulah hari keenam.

C.2.2 List of Top Score Set Terms

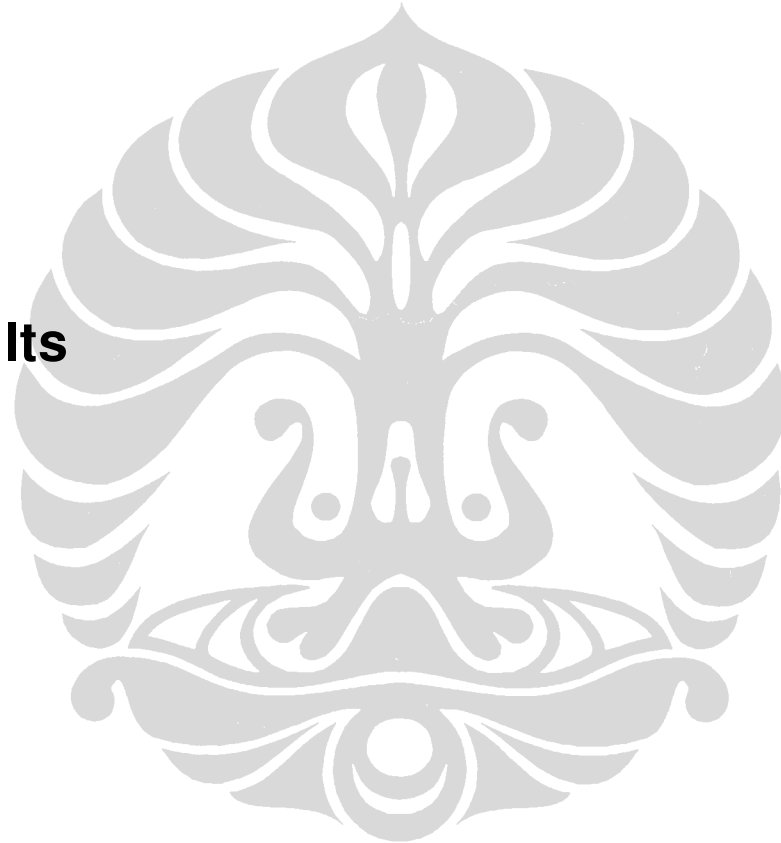
Table A.5 List Top Score Set Terms of P₁₀₀₀

No	Term	No	Term	No	Term	No	Term
1	earth	26	became	51	evil	76	truly
2	years	27	your	52	cain	77	time
3	went	28	wives	53	birds	78	taken
4	waters	29	sort of	54	beasts	79	sorrow
5	god	30	name	55	again	80	sin

6	came	31	gave	56	whom	81	seventh
7	you	32	everything	57	way	82	seed
8	seven	33	make	58	till	83	rest
9	every	34	great	59	those	84	place
10	life	35	flesh	60	things	85	our
11	end	36	wife	61	these	86	open
12	days	37	take	62	still	87	moving
13	man	38	sent	63	seen	88	kept
14	birth	39	saw	64	second	89	keep
15	said	40	male	65	rain	90	increased
16	living	41	living thing	66	my	91	image
17	them	42	heaven	67	give	92	generations
18	made	43	first	68	death	93	full
19	i	44	female	69	come	94	eyes
20	father	45	cattle	70	breath	95	deep
21	two	46	adam	71	bird	96	connection
22	lord	47	took	72	beast	97	young
23	him	48	son	73	air	98	work
24	face	49	put	74	year	99	without
25	day	50	goes	75	under	100	wind

Appendix B

Experiment Results



B.1 Bilingual Term Mapping Results

B.1.1 Parallel Corpus 1 for Random Set

- a. Including *stopwords* and no weighting

Table B.1 Bilingual Term Mapping Results for Random Set using Parallel Corpus 1 Including Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0000	0.0105	0.0000	0.0000	0.0526	0.0632	0.0421	0.0632	0.1053	0.1053	0.0947	0.0947
Top 10	0.0042	0.0042	0.0053	0.0053	0.0095	0.0179	0.0126	0.0116	0.0232	0.0179	0.0189	0.0189
Top 50	0.0040	0.0023	0.0032	0.0042	0.0057	0.0053	0.0053	0.0057	0.0076	0.0069	0.0076	0.0080
Top 100	0.0027	0.0019	0.0026	0.0027	0.0044	0.0036	0.0038	0.0038	0.0055	0.0051	0.0055	0.0052

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0000	0.0026	0.0000	0.0000	0.0206	0.0200	0.0148	0.0232	0.0394	0.0314	0.0270	0.0342
Top 10	0.0237	0.0158	0.0129	0.0125	0.0306	0.0473	0.0336	0.0334	0.0696	0.0406	0.0553	0.0609
Top 50	0.0635	0.0312	0.0402	0.0621	0.0852	0.0678	0.0642	0.0891	0.1064	0.0998	0.1103	0.1129
Top 100	0.0707	0.0455	0.0669	0.0727	0.1101	0.0910	0.1012	0.1008	0.1386	0.1300	0.1377	0.1324

b. Excluding *stopwords* and no weighting

Table B.2 Bilingual Term Mapping Results for Random Set using Parallel Corpus 1 Excluding Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0000	0.0000	0.0211	0.0105	0.0526	0.0737	0.0632	0.0421	0.0947	0.1053	0.0947	0.0632
Top 10	0.0042	0.0042	0.0084	0.0063	0.0095	0.0147	0.0147	0.0105	0.0221	0.0189	0.0168	0.0189
Top 50	0.0038	0.0025	0.0036	0.0042	0.0053	0.0044	0.0048	0.0053	0.0065	0.0063	0.0061	0.0065
Top 100	0.0027	0.0022	0.0025	0.0026	0.0041	0.0035	0.0038	0.0036	0.0048	0.0046	0.0047	0.0048

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0000	0.0000	0.0037	0.0026	0.0206	0.0215	0.0189	0.0171	0.0386	0.0309	0.0244	0.0288
Top 10	0.0237	0.0140	0.0247	0.0276	0.0306	0.0377	0.0414	0.0313	0.0672	0.0426	0.0556	0.0631
Top 50	0.0624	0.0351	0.0516	0.0616	0.0833	0.0543	0.0774	0.0853	0.1001	0.0867	0.0914	0.0991
Top 100	0.0717	0.0540	0.0666	0.0706	0.1063	0.0886	0.1045	0.0980	0.1313	0.1214	0.1292	0.1291

B.1.2 Parallel Corpus 1 for Top Score Set

a. Including *stopwords* and no weighting

Table B.3 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Including Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1591	0.0341	0.1250	0.1023	0.3750	0.2841	0.2841	0.3295	0.4659	0.3523	0.3295	0.3750
Top 10	0.0284	0.0159	0.0318	0.0250	0.0511	0.0489	0.0534	0.0557	0.0636	0.0670	0.0648	0.0625
Top 50	0.0109	0.0068	0.0105	0.0100	0.0150	0.0159	0.0155	0.0161	0.0170	0.0189	0.0184	0.0191
Top 100	0.0066	0.0047	0.0067	0.0074	0.0085	0.0099	0.0097	0.0095	0.0097	0.0108	0.0109	0.0114

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0639	0.0139	0.0301	0.0321	0.1795	0.1196	0.1221	0.1396	0.2055	0.1202	0.1196	0.1489
Top 10	0.1139	0.0465	0.0989	0.0992	0.2272	0.1722	0.2189	0.2348	0.2653	0.2374	0.2635	0.2433
Top 50	0.2116	0.0979	0.1602	0.1699	0.2950	0.2667	0.2871	0.3090	0.3002	0.3131	0.3146	0.3212
Top 100	0.2509	0.1338	0.2277	0.2500	0.3124	0.3131	0.3388	0.3366	0.3174	0.3340	0.3327	0.3437

b. Excluding *stopwords* and no weighting

Table B.4 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Excluding Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1591	0.0227	0.1023	0.1023	0.3750	0.2045	0.2614	0.3182	0.4659	0.2955	0.3750	0.3864
Top 10	0.0284	0.0136	0.0250	0.0250	0.0511	0.0409	0.0432	0.0466	0.0636	0.0511	0.0523	0.0591
Top 50	0.0109	0.0052	0.0077	0.0093	0.0150	0.0134	0.0127	0.0143	0.0170	0.0139	0.0157	0.0164
Top 100	0.0066	0.0033	0.0055	0.0063	0.0085	0.0076	0.0076	0.0083	0.0097	0.0081	0.0092	0.0097

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0639	0.0045	0.0498	0.0347	0.1795	0.0842	0.1318	0.1511	0.2055	0.1233	0.1791	0.1712
Top 10	0.1139	0.0431	0.0915	0.0981	0.2272	0.1640	0.1951	0.2065	0.2653	0.2141	0.2389	0.2536
Top 50	0.2116	0.0675	0.1091	0.1650	0.2950	0.2375	0.2510	0.2873	0.3002	0.2612	0.2934	0.2972
Top 100	0.2509	0.1004	0.1852	0.2308	0.3124	0.2722	0.2966	0.3097	0.3174	0.2872	0.3084	0.3174

c. Including *stopwords* and with TF-IDF weighting

Table B.5 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Including Stopwords and with TF-IDF Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1705	0.0682	0.0682	0.1136	0.3523	0.1818	0.2614	0.2955	0.4659	0.2614	0.3409	0.3977
Top 10	0.0284	0.0182	0.0227	0.0193	0.0568	0.0409	0.0466	0.0534	0.0682	0.0602	0.0625	0.0636
Top 50	0.0116	0.0070	0.0080	0.0082	0.0164	0.0136	0.0161	0.0164	0.0200	0.0175	0.0186	0.0186
Top 100	0.0074	0.0043	0.0051	0.0061	0.0100	0.0085	0.0092	0.0094	0.0117	0.0105	0.0110	0.0113

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0651	0.0176	0.0280	0.0343	0.1569	0.0623	0.1170	0.1299	0.1843	0.0949	0.1293	0.1350
Top 10	0.1055	0.0363	0.0643	0.0549	0.2415	0.1450	0.2026	0.2214	0.2724	0.2121	0.2381	0.2608
Top 50	0.2013	0.1060	0.0999	0.1393	0.3151	0.2343	0.2759	0.3000	0.3258	0.2859	0.3167	0.3153
Top 100	0.2661	0.1173	0.1427	0.2131	0.3425	0.2869	0.3125	0.3241	0.3469	0.3095	0.3352	0.3474

d. Excluding *stopwords* and with TF-IDF weighting

Table B.6 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Excluding Stopwords and with TF-IDF Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1591	0.0795	0.0568	0.0909	0.3750	0.1591	0.2386	0.2955	0.4659	0.2500	0.3409	0.3864
Top 10	0.0284	0.0148	0.0227	0.0205	0.0511	0.0386	0.0455	0.0466	0.0636	0.0534	0.0534	0.0580
Top 50	0.0109	0.0055	0.0068	0.0075	0.0150	0.0132	0.0141	0.0143	0.0170	0.0145	0.0157	0.0166
Top 100	0.0066	0.0039	0.0044	0.0056	0.0085	0.0073	0.0077	0.0084	0.0097	0.0084	0.0092	0.0092

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0639	0.0184	0.0159	0.0319	0.1795	0.0625	0.1238	0.1439	0.2055	0.1038	0.1405	0.1551
Top 10	0.1139	0.0305	0.0652	0.0700	0.2272	0.1420	0.2050	0.2063	0.2653	0.2171	0.2308	0.2505
Top 50	0.2116	0.0646	0.0904	0.1281	0.2950	0.2362	0.2691	0.2726	0.3002	0.2619	0.2914	0.2982
Top 100	0.2509	0.0931	0.1364	0.2016	0.3124	0.2589	0.2820	0.3125	0.3174	0.2891	0.3135	0.3167

e. Including *stopwords* and with Log Entropy weighting

Table B.7 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Including Stopwords and with Log Entropy Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.2841	0.0341	0.1477	0.2045	0.5000	0.3409	0.3977	0.5000	0.4773	0.3977	0.4432	0.4318
Top 10	0.0489	0.0148	0.0307	0.0352	0.0773	0.0614	0.0659	0.0739	0.0784	0.0670	0.0716	0.0750
Top 50	0.0159	0.0066	0.0111	0.0125	0.0205	0.0189	0.0209	0.0202	0.0223	0.0211	0.0223	0.0223
Top 100	0.0093	0.0047	0.0070	0.0083	0.0120	0.0110	0.0117	0.0116	0.0126	0.0126	0.0124	0.0126

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0887	0.0050	0.0476	0.0674	0.2050	0.1411	0.1607	0.1991	0.2000	0.1561	0.1784	0.1704
Top 10	0.1854	0.0379	0.1052	0.1217	0.2982	0.2263	0.2280	0.2686	0.2752	0.2253	0.2463	0.2771
Top 50	0.2780	0.1077	0.1810	0.2005	0.3427	0.2903	0.3482	0.3441	0.3544	0.3348	0.3507	0.3541
Top 100	0.3301	0.1622	0.2266	0.2690	0.3778	0.3343	0.3648	0.3636	0.3717	0.3670	0.3666	0.3707

f. Excluding *stopwords* and with Log Entropy weighting

Table B.8 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 1 Excluding Stopwords and with Log Entropy Weighting

Average Precision	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.2727	0.0682	0.1477	0.2500	0.5568	0.3295	0.4205	0.4886	0.5568	0.4091	0.4545	0.5455
Top 10	0.0545	0.0136	0.0341	0.0420	0.0727	0.0602	0.0693	0.0761	0.0795	0.0727	0.0795	0.0784
Top 50	0.0148	0.0052	0.0102	0.0118	0.0182	0.0164	0.0173	0.0175	0.0195	0.0189	0.0195	0.0191
Top 100	0.0081	0.0039	0.0068	0.0074	0.0098	0.0091	0.0098	0.0093	0.0101	0.0105	0.0105	0.0102

Average Recall	Size 100				Size 500				Size 1000			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0932	0.0140	0.0537	0.0783	0.2544	0.1308	0.1673	0.2087	0.2363	0.1602	0.1955	0.2245
Top 10	0.2138	0.0365	0.1260	0.1358	0.2885	0.2373	0.2820	0.2959	0.3012	0.2887	0.3027	0.3021
Top 50	0.2773	0.0858	0.1773	0.1937	0.3299	0.2869	0.3143	0.3167	0.3281	0.3190	0.3250	0.3227
Top 100	0.2958	0.1356	0.2227	0.2514	0.3411	0.3082	0.3282	0.3328	0.3408	0.3316	0.3366	0.3465

B.1.3 Random Baseline for Top Score Set of Parallel Corpus 1

Table B.9 Random Baseline Results for Top Score Set of Parallel Corpus 1

Average Precision	P100				P500				P1000			
	RNDM 1	RNDM 2	RNDM 3	Average	RNDM 1	RNDM 2	RNDM 3	Average	RNDM 1	RNDM 2	RNDM 3	Average
Top 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Top 10	0.0000	0.0000	0.0000	0.0000	0.0011	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
Top 50	0.0000	0.0002	0.0009	0.0004	0.0007	0.0002	0.0007	0.0005	0.0000	0.0000	0.0000	0.0000
Top 100	0.0003	0.0007	0.0006	0.0005	0.0006	0.0003	0.0006	0.0005	0.0002	0.0005	0.0002	0.0003

Average Recall	P100				P500				P1000			
	RNDM 1	RNDM 2	RNDM 3	Average	RNDM 1	RNDM 2	RNDM 3	Average	RNDM 1	RNDM 2	RNDM 3	Average
Top 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Top 10	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000
Top 50	0.0000	0.0008	0.0065	0.0024	0.0029	0.0005	0.0043	0.0026	0.0000	0.0000	0.0000	0.0000
Top 100	0.0125	0.0097	0.0072	0.0098	0.0094	0.0032	0.0061	0.0062	0.0067	0.0063	0.0064	0.0065

B.1.4 Parallel Corpus 2 for Top Score Set

a. Including *stopwords* and no weighting

Table B.10 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 2 Including Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000				Size 3500			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.2556	0.1333	0.1889	0.2000	0.2444	0.2111	0.1667	0.2222	0.4000	0.3667	0.3556	0.4000	0.4000	0.3556	0.3889	0.4111
Top 10	0.0456	0.0333	0.0367	0.0433	0.0633	0.0456	0.0556	0.0578	0.0733	0.0689	0.0644	0.0700	0.0933	0.0833	0.0922	0.0922
Top 50	0.0133	0.0111	0.0118	0.0124	0.0171	0.0136	0.0160	0.0173	0.0198	0.0173	0.0184	0.0189	0.0251	0.0260	0.0267	0.0253
Top 100	0.0074	0.0067	0.0069	0.0070	0.0093	0.0084	0.0093	0.0094	0.0109	0.0103	0.0106	0.0107	0.0144	0.0141	0.0141	0.0139

Average Recall	Size 100				Size 500				Size 1000				Size 3500			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0722	0.0447	0.0583	0.0660	0.0648	0.0528	0.0433	0.0517	0.1161	0.1079	0.0954	0.1104	0.1011	0.0920	0.0995	0.1022
Top 10	0.1435	0.0928	0.0942	0.1241	0.1590	0.1064	0.1263	0.1328	0.2124	0.2024	0.1875	0.1951	0.2212	0.1932	0.2136	0.2165
Top 50	0.1871	0.1320	0.1670	0.1814	0.1941	0.1552	0.1874	0.1969	0.2523	0.2317	0.2415	0.2455	0.2852	0.2912	0.2941	0.2855
Top 100	0.2043	0.1870	0.1997	0.1943	0.2072	0.1857	0.2100	0.2094	0.2646	0.2529	0.2661	0.2646	0.3057	0.3212	0.3121	0.3004

b. Excluding *stopwords* and no weighting

Table B.11 Bilingual Term Mapping Results for Top Score Set using Parallel Corpus 2 Excluding Stopwords and No Weighting

Average Precision	Size 100				Size 500				Size 1000				Size 3500			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.2333	0.1667	0.1889	0.2111	0.2556	0.1667	0.2000	0.2222	0.3667	0.3000	0.3111	0.3444	0.3667	0.3111	0.3333	0.3667
Top 10	0.0400	0.0300	0.0333	0.0400	0.0578	0.0422	0.0478	0.0489	0.0656	0.0578	0.0589	0.0622	0.0856	0.0722	0.0844	0.0856
Top 50	0.0109	0.0096	0.0104	0.0111	0.0147	0.0136	0.0147	0.0144	0.0173	0.0156	0.0162	0.0169	0.0236	0.0229	0.0236	0.0227
Top 100	0.0058	0.0056	0.0061	0.0062	0.0078	0.0076	0.0077	0.0076	0.0097	0.0091	0.0090	0.0092	0.0127	0.0129	0.0127	0.0128

Average Recall	Size 100				Size 500				Size 1000				Size 3500			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.0740	0.0555	0.0583	0.0629	0.0757	0.0398	0.0498	0.0622	0.1151	0.0905	0.0975	0.1029	0.1095	0.0870	0.0935	0.1095
Top 10	0.1262	0.0843	0.0989	0.1262	0.1388	0.0873	0.1152	0.1194	0.1850	0.1663	0.1724	0.1751	0.1996	0.1702	0.1997	0.1953
Top 50	0.1491	0.1238	0.1454	0.1559	0.1558	0.1532	0.1565	0.1551	0.2093	0.2016	0.2032	0.2083	0.2450	0.2479	0.2495	0.2394
Top 100	0.1525	0.1533	0.1618	0.1609	0.1592	0.1618	0.1616	0.1579	0.2207	0.2184	0.2127	0.2144	0.2572	0.2841	0.2656	0.2662

B.1.5 Text Alignment Granularity Experiment Results

- a. Including *stopwords* and no weighting

Table B.12 Text Alignment Granularity Experiment Results for Bible Verses and Chapters Including Stopwords and No Weighting

Average Precision	Verses				Chapters			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.4000	0.3556	0.3889	0.4111	0.2667	0.1778	0.2333	0.2778
Top 10	0.0933	0.0833	0.0922	0.0922	0.0589	0.0378	0.0478	0.0511
Top 50	0.0251	0.0260	0.0267	0.0253	0.0184	0.0127	0.0160	0.0176
Top 100	0.0144	0.0141	0.0141	0.0139	0.0112	0.0087	0.0094	0.0108

Average Recall	Verses				Chapters			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1011	0.0920	0.0995	0.1022	0.0693	0.0595	0.0690	0.0706
Top 10	0.2212	0.1932	0.2136	0.2165	0.1350	0.1010	0.1279	0.1322
Top 50	0.2852	0.2912	0.2941	0.2855	0.2359	0.1566	0.1880	0.2263
Top 100	0.3057	0.3212	0.3121	0.3004	0.2802	0.2211	0.2242	0.2763

b. Excluding *stopwords* and no weighting

Table B.13 Text alignment granularity Experiment Results for Bible Verses and Chapters Excluding Stopwords and No Weighting

Average Precision	Verses				Chapters			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.3667	0.3111	0.3333	0.3667	0.2444	0.1778	0.2222	0.2778
Top 10	0.0856	0.0722	0.0844	0.0856	0.0622	0.0367	0.0467	0.0556
Top 50	0.0236	0.0229	0.0236	0.0227	0.0173	0.0122	0.0151	0.0173
Top 100	0.0127	0.0129	0.0127	0.0128	0.0104	0.0076	0.0088	0.0099

Average Recall	Verses				Chapters			
	FREQ	10%	25%	50%	FREQ	10%	25%	50%
Top 1	0.1095	0.0870	0.0935	0.1095	0.0664	0.0565	0.0672	0.0711
Top 10	0.1996	0.1702	0.1997	0.1953	0.1546	0.0992	0.1205	0.1367
Top 50	0.2450	0.2479	0.2495	0.2394	0.2169	0.1464	0.1812	0.2154
Top 100	0.2572	0.2841	0.2656	0.2662	0.2544	0.1989	0.2131	0.2334

c. Comparison of Mapping Selection and Text alignment granularity

Table B.14 Comparison of Mapping Selection and Text Alignment Granularity

Mapping Selection	Verses		Chapters	
	P	R	P	R
Top 1	0.3852	0.0979	0.2296	0.0664
Top 10	0.0893	0.2078	0.0456	0.1204
Top 50	0.0260	0.2903	0.0154	0.1903
Top 100	0.0140	0.3112	0.0096	0.2405

d. Comparison of Rank Approximation and Text alignment granularity

Table B.15 Comparison of Rank Approximation and Text Alignment Granularity

Rank Approximation	Verses		Chapters	
	P	R	P	R
10%	0.1123	0.2109	0.0589	0.1299
25%	0.1220	0.2160	0.0749	0.1489
50%	0.1288	0.2144	0.0897	0.1703
100% (FREQ)	0.1277	0.2156	0.0862	0.1766

e. Comparison of Stopwords Usage and Text alignment granularity

Table B.16 Comparison of Stopwords Usage and Text Alignment Granularity

Stopwords Usage	Verses		Chapters	
	P	R	P	R
With Stopwords	0.1286	0.2268	0.0751	0.1544
No Stopwords	0.1134	0.2007	0.0740	0.1450

Appendix C

Full Paper in Proceedings of the Second International Malindo Workshop 2008, Malaysia

