

LAPORAN AKHIR PENELITIAN



**PENGEMBANGAN INFORMATION RETRIEVAL UNTUK ALKITAB
DALAM LAYANAN ALKITAB CERDAS**

TIM PENGUSUL

**Budi Susanto, S.Kom.,M.T
Gloria Virginia, S.Kom.,MAI.,Ph.D.
Umi Proboyekti, SKom, M.LIS.**

DUTA WACANA

Informatika

UNIVERSITAS KRISTEN DUTA WACANA

November 2020

DAFTAR ISI

DAFTAR ISI	i
DAFTAR TABEL	ii
DAFTAR GAMBAR	iii
RINGKASAN	1
BAB 1. PENDAHULUAN.....	2
1.1 Latar Belakang.....	2
1.2. Rumusan Masalah.....	2
1.3. Batasan Masalah	3
1.4. Rencana Target Capaian Tahunan	3
BAB 2. TINJAUAN PUSTAKA.....	5
BAB 3. TUJUAN DAN MANFAAT PENELITIAN	10
3.1. Road Map Penelitian.....	10
3.2. Tujuan Penelitian	10
3.3. Manfaat Penelitian	11
BAB 4. METODE PENELITIAN	12
BAB 5. HASIL DAN LUARAN YANG DICAPAI	13
5.1. Setup Apache Solr	13
5.2. Setup Core untuk Koleksi Alkitab	14
5.3. Konversi format Alkitab dari SQLite ke XML	17
5.4. Upload dan Indexing Dokumen SOLR.....	18
5.5. Instalasi SOLR di server.....	23
5.6. Aplikasi alkitabws	26
5.7. Instalasi Aplikasi alkitabws di server	31
5.8. Aplikasi Web Service berbasis RESTful Pencarian Alkitab Multi Bahasa	31
5.9. Uji Coba Web Service	40
Pengujian 1:	40
Pengujian 2:.....	42
Pengujian 3:.....	44
Pengujian 4:.....	46
BAB 6. RENCANA TAHAP BERIKUTNYA.....	51
BAB 7. KESIMPULAN DAN SARAN	52
7.1. Kesimpulan.....	52
7.2. Saran.....	52
DAFTAR PUSTAKA.....	53
LAMPIRAN	
A. DRAFT ARTIKEL	
B. BUKTI SUBMIT PENDAFTARAN HAK CIPTA	

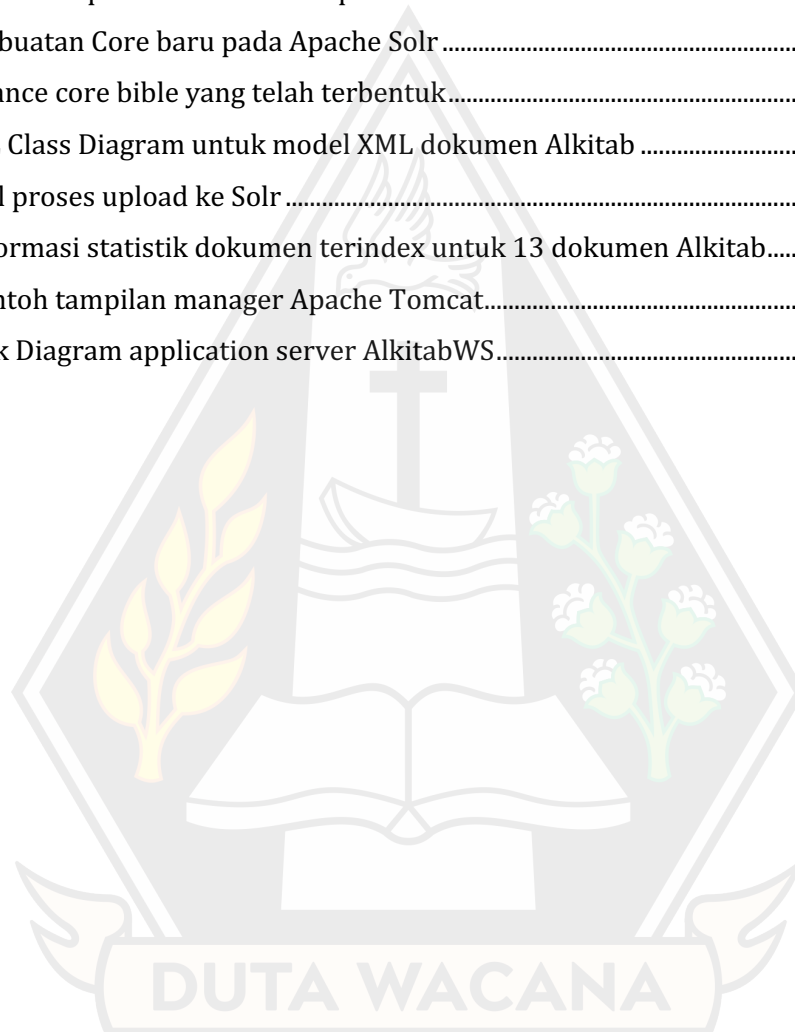
DAFTAR TABEL

Tabel 1. Daftar Alkitab yang Diindeks.....	3
Tabel 2. Rencana Target Capaian Tahunan.....	4
Tabel 3. Langkah-langkah penelitian.....	12
Tabel 4. Struktur data dokumen untuk representasi sebuah Ayat Alkitab	15
Tabel 5. Daftar field dinamis untuk multi bahasa dalam index Solr	15
Tabel 6. Pemetaan singkatan nama buku dalam Alkitab	26
Tabel 7. Daftar API Layanan Pencarian Alkitab Multi Bahasa.....	32



DAFTAR GAMBAR

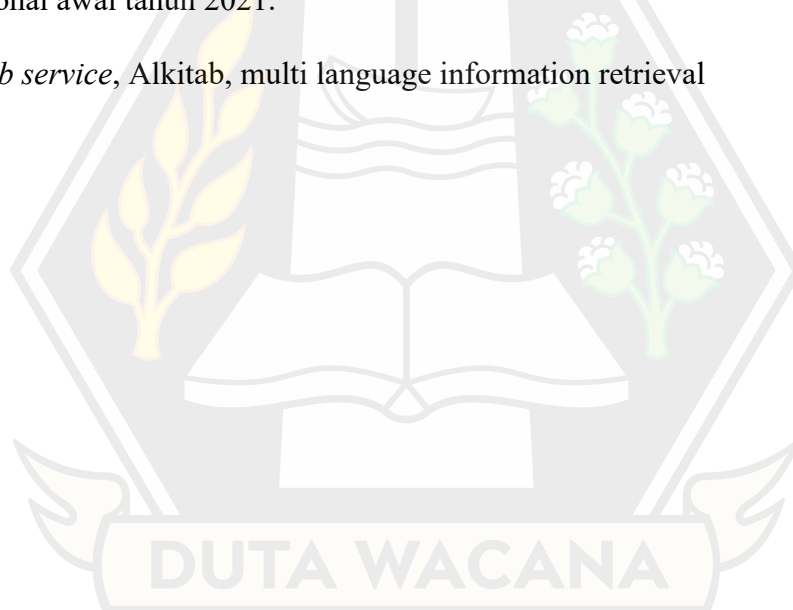
Gambar 1. Metode perbandingan dokumen dalam korpus multi bahasa	6
Gambar 2. Pendekatan kedua perbandingan dokumen-dokumen dalam korpus multi bahasa	7
Gambar 3. Diagram indexing dan querying untuk multi bahasa dengan multi index (sumber: (Grainger & Potter, 2014)	8
Gambar 4. Road map kegiatan penelitian Smart Bible	10
Gambar 5. Contoh tampilan home admin Apache Solr	14
Gambar 6. Pembuatan Core baru pada Apache Solr	16
Gambar 7. Instance core bible yang telah terbentuk.....	17
Gambar 8. UML Class Diagram untuk model XML dokumen Alkitab	18
Gambar 9. Hasil proses upload ke Solr	19
Gambar 10. Informasi statistik dokumen terindex untuk 13 dokumen Alkitab.....	19
Gambar 11. Contoh tampilan manager Apache Tomcat.....	31
Gambar 12. Blok Diagram application server AlkitabWS.....	40



RINGKASAN

Sebagai bagian dari pengembangan produk Smart Bible, pada penelitian ini akan dilakukan pengembangan platform pencarian korpus Alkitab yang memperhatikan multi bahasa. Tahap ini menjadi tahap awal paling penting. Platform temu kembali korpus Alkitab multi bahasa ini akan disediakan di atas protokol web service RESTful agar lebih terbuka digunakan dengan platform lainnya. Sebagai platform indexing dan retrieval yang membantu web service pencarian alkitab adalah Apache Solr. Pemanfaatan Solr telah memperhatikan prinsip penanganan pra proses terhadap dokumen Alkitab sesuai dengan bahasanya. Penelitian ini telah menghasilkan luaran produk konversi model dari dokumen Alkitab format XML ke Solr, demikian juga dari database SQLite ke XML. Sebagai satuan dokumen yang digunakan dalam Solr adalah ayat. Per ayat adalah sebuah dokumen dalam Solr. Dengan demikian hasil dari pengindeksan terhadap 13 bahasa dokumen Alkitab, diperoleh satuan dokumen ayat yang terindeks adalah sebanyak 402.595 dokumen. Dengan terindeksnya dokumen ayat tersebut, maka sistem pencarian Alkitab telah menyediakan infrastruktur terpenting dalam layanan web service pencarian Alkitab. Penelitian ini juga telah menghasilkan sebuah aplikasi web service yang memanfaatkan protokol RESTful untuk menyediakan antar muka pencarian teks Alkitab dan juga pengambilan teks Alkitab berdasar ekspresi rujukannya. Aplikasi ini telah diajukan untuk memperoleh Hak Cipta, dan telah dihasilkan draft artikel yang akan disubmit pada prosiding Nasional awal tahun 2021.

Kata kunci: *web service*, Alkitab, multi language information retrieval



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Sebagian besar pengguna, khususnya Kristiani, memiliki pengetahuan yang terbatas tentang Alkitab, dimana mereka tahu sebuah ayat atau cerita dalam Alkitab tetapi tidak dapat mengingat kata-kata yang tepat. Skenario pencarian yang khas ini, skema pencocokan secara leksikal mungkin tidak dapat memberikan kinerja yang baik. Rata-rata pengguna perangkat lunak pencarian Alkitab, kemampuan untuk mengambil ayat, cerita, atau bahkan topik dari Alkitab dengan menggunakan kata-kata mereka sendiri (yaitu bahasa alami) adalah pilihan yang sangat menarik. Produk pencarian Alkitab yang efektif perlu memungkinkan penggunaan bahasa alami untuk pencarian.

Di sisi lainnya bahwa Alkitab juga telah dipublikasikan dalam beberapa bahasa yaitu sebanyak 1.537 bahasa di seluruh dunia dan memiliki versi sebanyak 2.305 entri teks (Skounaki, 2019). Secara khusus untuk di Indonesia, menurut laporan yang dipublikasikan oleh Sabda.org (n.d.), terdapat 28 edisi terjemahan bahasa Indonesia, dan 47 terjemahan ke bahasa-bahasa daerah di Indonesia. Dalam sebuah Alkitab terdiri dari dua bagian, yaitu Perjanjian Lama (39 kitab) dan Perjanjian Baru (27 kitab) dengan jumlah total ayat adalah 31,102. Dengan beragamnya terjemahan Alkitab tersebut, ada suatu kebutuhan pencarian ayat atau cerita juga dapat dilakukan langsung pada kumpulan terjemahan tersebut (atau *cross-language information retrieval*).

Fungsi pencarian Alkitab ini menjadi satu hal mendasar untuk disediakan dalam sistem *smart bible*. yang menjadi rencana kegiatan penelitian selama 3 tahun mendatang yang diusulkan ini. *Smart bible* dalam kegiatan penelitian ini adalah sebuah sistem penyediaan layanan Alkitab *online* yang memperhatikan berbagai terjemahan, dan dikuatkan dengan onologi Alkitab terkait tema, tokoh, dan tempat. Oleh karena pencarian Alkitab menjadi fungsi dasar dalam sistem *smart bible*, maka fungsi pencarian Alkitab harus dapat menerapkan protokol RESTful.

1.2. Rumusan Masalah

Kegiatan penelitian tahun pertama ini direncanakan fokus untuk penerapan Apache SOLR untuk mendukung pencarian dengan mekanisme temu kembali (*information retrieval*) untuk *corpus* Alkitab dalam beberapa terjemahan penuh. Tantangan lain yang perlu diselesaikan dalam kegiatan penelitian tahun pertama ini adalah penerapan sistem temu

kembali ini harus dapat diakses melalui protokol RESTful yang juga memperhatikan keamanan aksesnya.

1.3. Batasan Masalah

Publikasi Alkitab terjemahan yang akan digunakan dalam pengujian sistem pencarian ini seperti ditunjukkan pada Tabel 1. Adapun format alkitab yang dijadikan dasar adalah format XML yang digunakan oleh Thiago Bodruk¹. Oleh karena dari daftar Alkitab yang tersedia oleh Thiago Bodruk tidak terdapat Alkitab Indonesia, maka data Alkitab Terjemahan Baru Indonesia diambil dari database Alkitab yang digunakan oleh aplikasi My Bible²³.

Tabel 1. Daftar Alkitab yang Diindeks

No.	Lang	Bible Version
1	Korean	Korean Version
2	France	Le Bible de l'Épée
3	German	Schlachter
4	Greek	Modern Greek
5	Finnish	Finnish Bible
6	Spanish	Reina Valera
7	Portuguese	Almeida Revisada Imprensa Bíblica
8	Chinese	New Chinese Version
9	Romanian	Versiunea Dumitru Cornilescu
10	Arabic	The Arabic Bible
11	English	KJV
12	Russian	Синодальный перевод
13	Indonesia	Terjemahan Baru Tahun 1974

1.4. Rencana Target Capaian Tahunan

Penelitian 3 tahun yang direncanakan mulai tahun 2020 ini ditargetkan dapat menghasilkan sebuah produk akhir berupa sistem yang memberikan layanan berbasis web service untuk pencarian Alkitab. Tabel 2 memperlihatkan rencana keluaran yang direncanakan, dimana TS adalah tahun 2020.

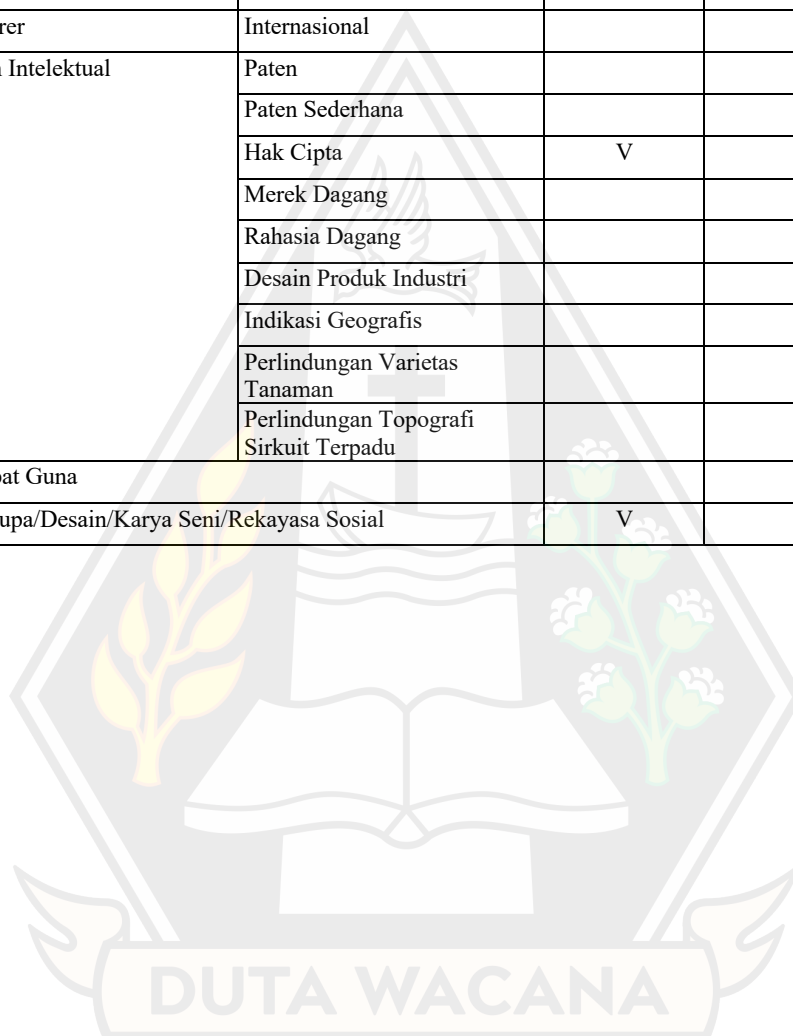
¹ <https://github.com/thiagobodruk/bible>

² <https://mybible.com/>

³ https://www.ph4.org/_dl.php?back=bbl&a=TB&b=mybible&c

Tabel 2. Rencana Target Capaian Tahunan

No	Jenis Luaran				Indikator Capaian
	Kategori	Sub Kategori	Wajib	Tambahan	TS
1	Artikel Ilmiah dimuat di jurnal	Internasional bereputasi		V	Submitted
		Nasional Terakreditasi			
2	Artikel Ilmiah dimuat di prosiding	Internasional terindeks		V	Published
		Nasional			
3	<i>Invited speaker</i> dalam temu ilmiah	Internasional			
		Nasional			
4	Visiting Lecturer	Internasional			
5	Hak Kekayaan Intelektual	Paten			
		Paten Sederhana			
		Hak Cipta	V		granted
		Merek Dagang			
		Rahasia Dagang			
		Desain Produk Industri			
		Indikasi Geografis			
		Perlindungan Varietas Tanaman			
		Perlindungan Topografi Sirkuit Terpadu			
6	Teknologi Tepat Guna				
7	Model/Purwarupa/Desain/Karya Seni/Rekayasa Sosial		V		Produk



BAB 7. KESIMPULAN DAN SARAN

7.1. Kesimpulan

Penelitian ini telah berhasil mendefinisikan struktu dokumen dalam Apache Solr dan juga telah melakukan index sebanyak 402.595 dokumen ayat dari 13 dokumen Alkitab dengan berbagai bahasa. Bahasa yang dipilih tidak hanya berbasis pada ANSI ASCII namun juga telah memperhatikan UNICODE untuk beberapa bahasa seperti Arab, Korea, Cina, atau Perancis. Untuk dapat melakukan pengindeksan ini, penelitian ini telah menghasilkan modul konversi otomatis dari format XML ke dalam dokumen Solr. Modul ini menjadi sangat penting sekali mengingat struktur dokumen ayat Alkitab yang digunakan juga tidak mengikuti definisi konfigurasi baku. Hal paling mendasar terkait hal ini adalah karena dokumen yang diindex harus memperhatikan bahasa-bahasa yang berbeda.

Hasil dari konfigurasi mesin pencari Apache SOLR untuk mendukung pencarian multi bahasa dapat terbukti menghasilkan pencarian sesuai dengan kebutuhan. Demikian hasil pengujian terhadap aplikasi web service yang memberikan fungsi sebagai gerbang untuk mesin pencari Apache Solr. Aplikasi web service berbasis RESTful untuk mendukung Pencarian Teks Alkitab Multi Bahasa telah menunjukkan keberhasilan eksekusi fungsi, baik fungsi otorisasi yang memanfaatkan protokol JWT dan juga fungsi pencarian serta pengambilan teks Alkitab berdasar rujukan perikop.

7.2. Saran

Dengan selesainya tahap pertama pengembangan aplikasi web service untuk pencarian Alkitab Multi Bahasa, dapat membantu siapa saja untuk dapat memanfaatkan dalam penyediaan aplikasi-aplikasi yang berbasis Alkitab. Hal ini penting untuk terus didorong agar juga dapat menguatkan pelayanan misi penyebaran Firman Tuhan di dunia. Walaupun terdapat banyak layanan penyedia teks Alkitab, namun demikian dengan dasar hasil penelitian tahun pertama ini, pada tahun selanjutnya dapat dilakukan pengayaan konten dan proses yang lebih cerdas. Untuk itulah perlu dilakukan pengembangan beberapa contoh aplikasi yang dapat memanfaatkan layanan pencarian teks Alkitab multi bahasa ini.

DAFTAR PUSTAKA

- Skounaki, S. (2019, Agustus 15). *August 2019 Content Update*. Retrieved from The Digital Bible Library: <https://thedigitalbiblelibrary.org/2019/08/15/august-2019-content-update/#more-4128>
- Sabda.org. (n.d.). *Versi Alkitab*. Retrieved from Sabda: <https://alkitab.mobi/tb/versions/>
- Resnik, P., Olsen, M. B., & Diab, M. (1999). The Bible as a Parallel Corpus: Annotating the "Book of 2000 Tongues". *Computers and the Humanities*, 33(1/2), 129-153.
- Roorda, D., Kalkman, G., Naaijer, M., & Cranenburgh, A. v. (2014, Desember 4). LAF-Fabric: a data analysis tool for Linguistic Annotation Framework with an application to the Hebrew Bible. *Computational Linguistics in the Netherlands Journal*, 4, 105 - 109.
- Christodouloupoulos, C., & Steedman, M. (2015). A massively parallel corpus: the Bible in 100 languages. *Language Resources and Evaluation*, 49, 375-395.
- Berry, M. W., & Young, P. G. (1995). Using Latent Semantic Indexing for Multilanguage Information Retrieval. *Computers and the Humanities*, 29(6), 413-429.
- Chew, P. A., Verzi, S. J., Bauer, T. L., & McClain, J. T. (2006). Evaluation of the Bible as a Resource for Cross-Language Information Retrieval. *Proceedings of the Workshop on Multilingual Language Resources and Interoperability* (pp. 68-74). Sydney, Australia: Association for Computational Linguistics.
- Airio, E. (2008). Who benefits from CLIR in web retrieval? *Journal of Documentation*, 64(5), 760-778.
- SOLR, A. (n.d.). *Apache Solr Reference Guide*. Retrieved from Language Analysis: https://lucene.apache.org/solr/guide/8_6/language-analysis.html
- Grainger, T., & Potter, T. (2014). *Solr in Action*. Shelter Island, NY: Manning Publications Co.
- Kumar, J. (2015). *Apache Solr Search Pattern*. Birmingham, UK: Packt Publishing.

DUTA WACANA