

Pemodelan Ontologi untuk Klasifikasi Dokumen Informatika

by Ulfa Amalia Nur Fatimah

Submission date: 19-Nov-2019 08:44PM (UTC+0700)

Submission ID: 1216453423

File name: pemodelan_ontologi_utk_klasifikasi_dokumen_informatika.docx (272.76K)

Word count: 2351

Character count: 15254

Pemodelan Ontologi Untuk Klasifikasi Dokumen Informatika

Abstrak—Teknologi semantik rupanya telah digunakan dalam beberapa tugas di dunia *computer science*, salah satunya adalah klasifikasi dokumen. Banyak cara yang dilakukan klasifikasi dokumen menggunakan teknologi semantik, salah satunya adalah memanfaatkan ontologi, baik itu hanya ontologi saja ataupun digabung dengan metode klasifikasi tradisional (*hybrid method*). Dalam masalah klasifikasi dokumen proposal Informatika UII ini menggunakan domain *computer science* sebagai *superclass*-nya. Subkelas yang ada terdapat enam buah, di antaranya adalah Komputasi dan Sistem Cerdas, Multimedia Visi Komputer, Informatika Medis, Sistem Informasi, Jaringan dan Keamanan Komputer, dan Rekayasa Perangkat Lunak dengan bersumber pada *digital library* dari *Association of Computing Machinery* (ACM) yang disesuaikan dengan konsentrasi-konsentrasi yang ada di Informatika UII. Pembuatan ontologi ini menggunakan *tool Protege* karena mudah digunakan. Hasil ontologi yang diperoleh dapat dimanfaatkan untuk klasifikasi dokumen.

Kata kunci—ontologi; taksonomi; semantik; computer science;

I. PENDAHULUAN 17

Istilah “Revolusi Industri 4.0” merupakan istilah yang sering kita dengar akhir-akhir ini. Secara tidak sadar, pemanfaatan teknologi, lambat laun, mendominasi segala aspek kehidupan manusia. Dengan kemampuan teknologi informasi yang cepat dan mudah diakses ini manusia dapat memperoleh segala informasi dan layanan yang diinginkan. Salah satu contoh pemanfaatan teknologi informasi yang populer adalah pemanfaatan portal web.

Portal web merupakan situs web yang dirancang khusus sebagai titik akses tunggal untuk pencarian informasi [1]. Kita dapat melakukan pencarian, pengelolaan, pengintegrasian informasi, dan lainnya. Portal web mengumpulkan banyak informasi dari sekian banyak sumber informasi lalu dikumpulkan dalam satu tempat. Di ruang lingkup organisasi, portal web dapat disesuaikan dengan peran (*role*) dari pengguna yang didapatkan [1].

Web sudah mengalami perkembangan signifikan sampai saat ini dan fungsinya juga semakin berkembang mengikuti pertumbuhan kebutuhan manusia. Perkembangan dari Web 1.0, Web 2.0, hingga Web 3.0 seperti sekarang ini.

1. Web 1.0 (1990-2000)

Ditemukan oleh Tim Berners-Lee, penemu *World Wide Web*. Web tersebut hanya dapat dibaca (*read-only platform*) dan statis [2]. Tujuan utama adanya Web ini adalah mempublikasikan informasi seperti iklan bisnis dan para konsumen dapat menghubungi dari iklan tersebut [3]. Karena hanya bertujuan menampilkan informasi, web ini sangat tidak interaktif dan bahkan *link* yang tersedia juga terlampau lemah [3]. Teknologi inti dari web ini diantaranya adalah

HTTP, HTML, dan URI [3] dan contoh web yang tersedia pada era Web 1.0 adalah *Netscape Navigator* [2].

2. Web 2.0 (2000-2010)

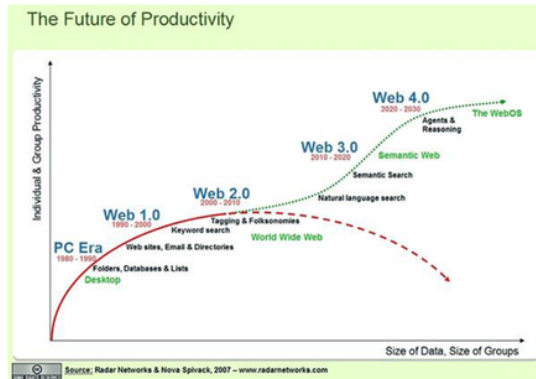
Dari era Web 1.0 yang merupakan *read-only platform* yang menyebabkan para pengguna sangat pasif terhadap web, maka dikembangkanlah versi Web 2.0. Istilah Web 2.0 ini diprakarsai oleh Dale Dougherty, Wakil Direktur dari O'Reilly Media [3]. Di era ini, para pengguna tidak hanya dapat membaca informasi, tetapi juga dapat menyediakan informasi ke web melalui internet yang sekaligus juga dapat membaginya ke sesama pengguna [2].

3. Web 3.0 (2010-sekarang)

Web 3.0 juga dikenal sebagai *Semantic Web*. Ide ini berasal dari Tim Berners-Lee [3]. *Semantic* (atau Semantik dalam bahasa Indonesia) berarti mempelajari tentang makna [2]. *Semantic Web* atau Web Semantik merupakan web yang dapat menunjukkan sesuatu dengan pendekatan yang dapat dimengerti oleh komputer dan bertujuan utama agar web dapat dibaca oleh mesin dan bukan hanya manusia [3]. Web Semantik juga berdasarkan konten digital dengan kumpulan kosa kata standar yang menyediakan semantik agar dapat dimengerti oleh komputer [2].

4. Web 4.0 (20...-20...)

Web di era ini masih dalam perkembangan dan belum diketahui akan seperti apa web tersebut. Namun, Web 4.0 juga disebut dengan *sybiotic web*, dimana hubungan antara manusia dan mesin di dalam simbiosis yang tentu saja saling menguntungkan [3]. Web ini juga pasti akan lebih banyak memanfaatkan kecerdasan buatan yang nantinya bisa disebut juga sebagai *intelligent web* [3].



GAMBAR 1. EVOLUSI DARI WEB

Sumber: <http://www.mokabyte.it/>

Pemanfaatan web, baik web yang sederhana maupun canggih, sudah dimanfaatkan dalam berbagai domain. Di sektor pendidikan, khususnya, pengelolaan dokumen berbasis web sudah banyak dilakukan di perguruan tinggi-perguruan tinggi di dunia, termasuk di Indonesia. Portal web tersebut memfasilitasi mahasiswa dan dosen untuk mempermudah mengelola dokumen secara elektronik, meliputi mengklasifikasikan atau mengelompokkan dokumen-dokumen sesuai dengan jurusan maupun konsentrasi penelitian. Di samping itu, pencarian dokumen lebih mudah dilakukan dengan menggunakan portal web.

Program Studi (Prodi) Informatika Fakultas Teknologi Industri Universitas Islam Indonesia, khususnya Program Sarjana, adalah salah satu yang menerapkan penggunaan portal web untuk mengelola dokumen, khususnya untuk tugas akhir mahasiswa. Nama portal tersebut adalah Kerja Praktik dan Tugas Akhir Fakultas Teknologi Industri Universitas Islam Indonesia (KPTA FTI UII). Proposal yang diajukan tersebut diunggah sebagai persyaratan tugas akhir mahasiswa. Penggunaan portal ini dinilai cukup membantu dan mudah digunakan bagi mahasiswa dan dosen pembimbing. Mahasiswa yang sudah memiliki akun KPTA FTI UII akan masuk dan memilih konsentrasi yang telah ditetapkan, mengunggah proposalnya, kemudian diseleksi oleh para dosen pengampu konsentrasi. Namun, yang menjadi masalah adalah sering kali ditemukan banyak mahasiswa membuat proposal yang tidak sesuai dengan konsentrasi yang dipilihnya. Selain itu, saat para dosen menentukan apakah proposal diterima atau tidak, para dosen harus memeriksa secara teliti proposal tersebut apakah sesuai dengan konsentrasi atau tidak dan itu membutuhkan waktu yang tidak sebentar. Jika tidak diterima, kemungkinan proposal tersebut dapat ditolak dan dapat menghambat pengerjaan skripsi mahasiswa karena mahasiswa harus menunggu keputusan dosen dan mengunggah proposal yang baru. Dilihat dari permasalahan tersebut, maka perlu adanya perbaikan model KPTA FTI UII agar dapat bekerja lebih efisien dan hemat waktu.

Pemodelan baru untuk klasifikasi dokumen dengan pendekatan web semantik merupakan terobosan baru karena di dalam prosesnya, komputer dapat memahami basis pengetahuan yang berhubungan dengan dokumen tersebut. Salah satu teknologi semantik yang populer adalah ontologi. Ontologi sendiri merupakan deskripsi dari konsep dan relasi atau spesifikasi yang nyata dari konseptualisasi [2]. Definisi ini konsisten dengan kegunaannya yang merupakan satu set definisi konsep [2]. Peran ontologi sangat penting dalam ilmu pengetahuan modern karena ontologi dapat memberikan standar yang formal dalam penyusunan basis pengetahuan [4].

Makalah ini disusun dengan urutan sebagai berikut. Bagian 2 membahas mengenai penelitian serupa yang pernah dilakukan sebelumnya. Bagian 3 membahas mengenai taksonomi dan perancangan ontologi. Bagian 4 menjelaskan mengenai hasil ontologi yang telah dibuat. Kesimpulan dari penelitian ini berada di Bagian 5.

II. PENELITIAN TERKAIT

Penelitian mengenai klasifikasi dokumen dengan memanfaatkan ontologi sudah terlaksana dari beberapa tahun sebelumnya. Penelitian dari Qazi [5] yang menerapkan ontologi domain olahraga. Dalam penelitiannya, Qazi menerapkan pembobotan kata berdasarkan ontologi yang telah dibuatnya. Selain itu, ia juga menerapkan metode klasifikasi tradisional (*Multinomial Naive Bayes*, *Decision Tree*, *Nearest Centroid*, *KNN*) untuk membandingkan hasilnya. Penelitiannya membuktikan bahwa menerapkan pembobotan kata berdasarkan ontologi kata menambah akurasi klasifikasi dokumen.

Penelitian serupa juga dilakukan oleh Zhou [6]. Dalam penelitiannya, ia menciptakan ontologi dan mengklasifikasikannya dengan algoritma *Unsupervised Deep Learning*. Berdasarkan hasil pengujiannya menunjukkan bahwa *recall* dan *precision* mengalami kenaikan yang signifikan dibandingkan dengan klasifikasi tanpa implementasi ontologi di dalamnya.

Fraihat [7] juga menerapkan ontologi konsep untuk penelitiannya dengan bantuan algoritma *Support Vector Machine* (SVM) untuk klasifikasi dokumennya. Hasilnya juga menunjukkan penerapan ontologi menghasilkan kenaikan akurasi yang signifikan dibandingkan dengan hanya menggunakan SVM.

Dari hasil penelitian sebelumnya menunjukkan bahwa penggunaan ontologi dalam hal klasifikasi dokumen dapat meningkatkan akurasi karena salah satu keuntungan ontologi adalah dapat memberikan pemahaman atau definisi bersama tentang konsep kunci tertentu di dalam sebuah domain [8]. Dengan bahasa deskripsi ontologi yang dapat menyandikan pengetahuan dan semantiknya seperti *Resource Definition Framework* (RDF) dan *Web Ontology Language* (OWL), komputer dapat memahami domain yang kita ajukan [8].

III. TAKSONOMI DAN ONTOLOGI

Dasar dari ontologi adalah taksonomi. Taksonomi berasal dari bahasa Yunani *Taxis* dan *Nomos* yang artinya 'mengelompokkan' dan 'peraturan atau hukum'. Taksonomi

merupakan klasifikasi dari benda dalam bentuk hirarki (tingkatan) [2]. Pada awalnya, taksonomi hanya digunakan pada klasifikasi makhluk hidup. Namun, sekarang taksonomi dapat digunakan untuk hal-hal lainnya, merujuk pada pengelompokan benda atau konsep.

Selain taksonomi, terdapat pula istilah *thesaurus* yang dimana ini merujuk ke perluasan dari taksonomi. *Thesaurus* berarti kumpulan kata yang dikelompokkan sesuai dengan kesamaan makna (baik sinonim maupun antonim).

Antara taksonomi dan ontologi keduanya merupakan hal yang berbeda. Taksonomi digunakan untuk mengorganisir pengetahuan, sedangkan ontologi merupakan representasi dari ontologi. Taksonomi dirasa lebih lemah untuk merepresentasikan pengetahuan [8] dan ontologi menyiratkan perspektif yang lebih luas dari taksonomi [2]. Namun, secara teori terdapat beberapa perbedaan jika dicermati. Menurut [2], berikut ini merupakan tabel perbedaan antara taksonomi dan ontologi:

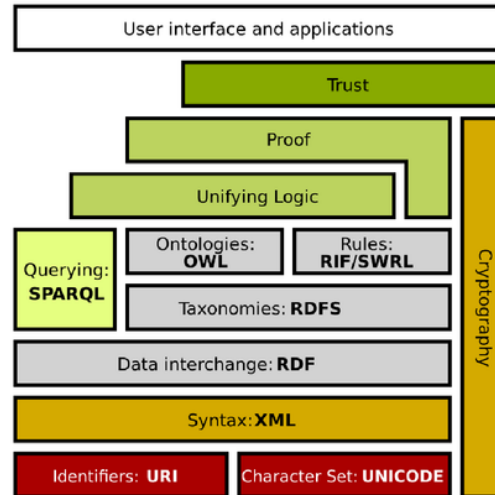
TABEL 1. PERBEDAAN TAKSONOMI DAN ONTOLOGI

| Taksonomi | Ontologi |
|--|---|
| Didefinisikan sebagai pengelompokan elemen secara urut yang mengindikasikan hubungan yang alami. | Dapat didefinisikan sebagai spesifikasi yang eksplisit dari konseptualisasi. |
| Secara umum tidak memasukkan kata kunci yang deskriptif untuk masing-masing elemen. | Secara umum memasukkan istilah yang deskriptif untuk mengelompokkan kosakata dari konten-konten <i>node</i> . |
| Taksonomi lebih sederhana dan berkembang untuk mengatur subjek dengan cara tertentu. | Ontologi dapat merepresentasikan interkoneksi yang sangat rumit. |
| Taksonomi cenderung cukup toleran tentang jenis hubungan yang ada diantara <i>parents node</i> dan <i>child node</i> di pohon klasifikasi. | Relasi di ontologi mengikuti tafsiran subjektif sehingga memicu adanya inferensi. |

Karena taksonomi juga merupakan peta pengetahuan (*knowledge map*), taksonomi harus dapat dipahami oleh pengguna, dalam artian bahwa taksonomi tersebut dapat menggambarkan hubungan antara istilah-istilah di dalamnya dan dapat dikembangkan menjadi ontologi.

Saat kita mengembangkan sebuah ontologi, nantinya ontologi tersebut akan menghasilkan *Web Ontology Language* (OWL) [9]. Menurut *World Wide Web Consortium* (W3C), OWL juga merepresentasikan pengetahuan yang kompleks dan kaya mengenai sesuatu, suatu kelompok, dan hubungan antara

sesuatu. Beberapa mengira bahwa OWL dan *Resource Development Framework* (RDF) adalah dua hal yang sama, akan tetapi keduanya berbeda. RDF merupakan standar untuk pertukaran data di web yang menggunakan *Uniform Resource Identifier* (URI) untuk memberi nama hubungan antara sesuatu dan begitu juga dengan kedua ujung tautan (biasa disebut *triple*). Antara OWL dan RDF dapat dibedakan seperti pada gambar bagian-bagian pada web semantik, menurut W3C, di bawah ini.



GAMBAR 2 LAPISAN WEB SEMANTIK MENURUT W3C

12

Sumber: https://en.wikipedia.org/wiki/Semantic_Web_Stack

Menurut [10], terdapat beberapa langkah untuk mengembangkan ontologi, diantaranya:

1. Mendefinisikan domain atau konsep
2. Mengatur kelas dalam hierarki (*subclass-superclass*) taksonomi
3. Mendefinisikan *slot* (biasa disebut dengan atribut properti)
4. Mendefinisikan *facet* (batasan) pada *slot*

Masing-masing langkah pengembangan ontologi dijelaskan di bawah ini dan dilanjutkan di bagian selanjutnya.

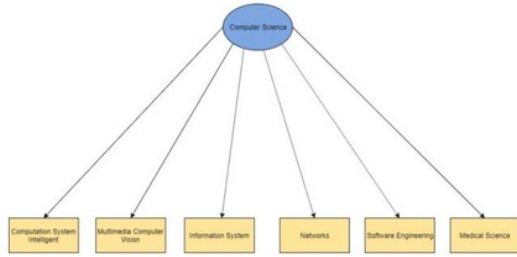
Mendefinisikan domain

Dalam penelitian ini, taksonomi dan ontologi yang dibangun diatas domain informatika (*computer science*). Selanjutnya, beberapa sub kelas yang berkaitan berisi tentang konsentrasi-konsentrasi yang ada di Informatika UII, yaitu Komputasi dan Sistem Cerdas, Multimedia Visi Komputer, Informatika Medis, Sistem Informasi, Jaringan dan Keamanan Komputer, dan Rekayasa Perangkat Lunak.

Patokan sub kelas yang digunakan berasal dari *digital library* dari *Association for Computing Machinery* (ACM)

dengan menyesuaikan konsentrasi-konsentrasi yang ada di Informatika UII.

Mengatur Kelas dalam Bentuk Taksonomi



GAMBAR 3. TAKSONOMI DARI DOMAIN COMPUTER SCIENCE

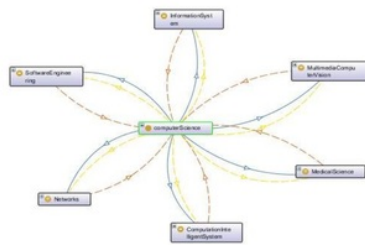
Merujuk dari gambar taksonomi di atas, *computer science* merupakan kelas utama (*superclass*) yang memiliki enam sub kelas. Sub kelas ditunjukkan dengan anak panah ke bawah. Dari keenam sub kelas yang tercantum masih terdapat sub sub kelas di bawahnya. Namun, karena adanya keterbatasan tempat gambar yang dimuat hanya sebagian.

Pembentukan taksonomi menggunakan dengan pendekatan *top-down*, dimana proses pengembangan dimulai dari definisi yang paling umum di sebuah domain dengan menyesuaikan konsentrasi-konsentrasi yang ada di Informatika UII.

Dari taksonomi diatas, maka akan dikembangkan sebuah ontologi dengan domain *computer science* di bagian selanjutnya.

IV. PENGEMBANGAN ONTOLOGI

Ontologi yang dibuat berdasarkan taksonomi yang dirancang sebelumnya. *Tool* yang digunakan untuk membuat ontologi adalah *Protege 5.5.0* yang dikembangkan oleh Stanford University karena terbuka [11], mudah digunakan dan mudah mendapat gambaran ontologi secara visualnya.



GAMBAR 4. ILUSTRASI ONTOLOGI MENGGUNAKAN PLUGIN ONTOGRAF

Mendefinisikan Slot

Slot atau biasa disebut atribut properti mendeskripsikan fitur-fitur dari konsep [10]. Tipe slot yang digunakan adalah hubungan (*relationship*) antara anggota individu kelas dengan yang lainnya.

Ontologi di atas merupakan gambaran dari *plugin OntoGraf* di dalam *Protege*. Ontologi yang telah dirancang memiliki hubungan *hasBranch* (panah kuning) dan *isPartOf* (panah coklat). Hubungan *hasBranch* berasal dari *domain computer science* menuju *range* keenam sub kelas yang tersedia. Sedangkan *isPartOf* adalah tipe properti *inverse functional* atau kebalikan dari panah *hasBranch* yang merupakan tipe properti *functional*. Sedangkan panah biru menunjukkan *subclass* dari *superclass computer science*. Sekali lagi yang ditunjukkan di dalam gambar merupakan sebagian kecil dari ontologi yang dibuat.



GAMBAR 5. ILUSTRASI ONTOLOGI MENGGUNAKAN PLUGIN OWLVIZ

Gambar ontologi di atas juga merupakan ontologi yang sama dengan memanfaatkan *plugin OWLViz*. Gambar ini menunjukkan *is-a relation* dimana itu menunjukkan *inheritance* dan sub kelas dari kelas sebelumnya. Terdapat juga kelas yang berbahasa Indonesia (misal, terdapat subkelas "Sistem Cerdas" dan seterusnya) yang disamakan (*equivalent*) dengan subkelas versi bahasa Inggris (misal, subkelas "Computational Intelligent System" dan seterusnya) dengan harapan untuk menyamakannya dengan versi Inggris. Di dalam gambar ontologi tersebut juga dicontohkan beberapa sub sub kelas di dalam subkelas *computational intelligent system* yang juga berelasi *is-a*.

Mendefinisikan Facet

Facet memiliki dua tipe, yaitu *slot cardinality* dan *slot value type*. Tipe yang digunakan disini adalah tipe *slot value type*. Sebuah tipe facet disini dapat menggambarkan tipe dari nilai yang mengisi *slot* [10]. Tipe nilai umum yang digunakan dalam kasus ini adalah *string*.

KESIMPULAN

Ontologi merupakan perpanjangan dari taksonomi dan taksonomi juga merupakan salah satu tahap pembuatan ontologi. Ontologi domain informatika (*computer science*) telah dikembangkan dengan tujuan sebagai alat untuk mengklasifikasikan dokumen proposal di dalam ruang lingkup

Informatika UII. Dengan memanfaatkan *Protege* sebagai *tool* untuk membuat ontologi secara visual, ontologi dengan domain *computer science* telah dikembangkan dan selanjutnya dapat dimanfaatkan untuk klasifikasi dokumen berbasis semantik.

REFERENSI

- [1] E. Rosario, J. Rosario, M. Nieva, T. Tan, and M. Tangkeko, "CollaborateIT: A CCS IT Thesis Portal with Electronic Document Management System," vol. 4, pp. 4–9, 2016.
- [2] D. Nandini, *Semantic Web and Ontology*. 2014.
- [3] S. Aghaei, M. A. Nematbakhsh, and H. K. Farsani, "E VOLUTION OF THE WORLD WIDE WEB: FROM," vol. 3, no. 1, pp. 1–10, 2012.
- [4] E. C. Pembrani, "Pembelajaran Relasi Non-Taksonomi Dari Dokumen Web."
- [5] A. Qazi and R. H. Goudar, "ScienceDirect An Ontology-based Term Weighting Technique for Web Document Categorization," *Procedia Comput. Sci.*, vol. 133, pp. 75–81, 2018.
- [6] P. Zhou, S. M. Asce, N. El-gohary, and A. M. Asce, "Ontology-Based Multilabel Text Classification of Construction Regulatory Documents," pp. 1–13, 2015.
- [7] M. M. Al-tahrawi and S. N. Al-khatib, "Ig Ht S Y Ig Ht S," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 27, no. 4, pp. 437–449, 2015.
- [8] Yu, Liyang, *A Developer's Guide to the Semantic Web*. Berlin: Springer, 2014.
- [9] A. Nugroho, "Membangun ontologi jurnal menggunakan protégé (."
- [10] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," pp. 1–25, 2000.
- [11] D. L. Rubin, N. F. Noy, and M. A. Musen, "Protégé: A Tool for Managing and Using Terminology in Radiology Applications," vol. 20, no. August, pp. 34–46, 2007.

Pemodelan Ontologi untuk Klasifikasi Dokumen Informatika

ORIGINALITY REPORT

10%

SIMILARITY INDEX

8%

INTERNET SOURCES

5%

PUBLICATIONS

6%

STUDENT PAPERS

PRIMARY SOURCES

| | | |
|---|---|----|
| 1 | tel.archives-ouvertes.fr Internet Source | 1% |
| 2 | Submitted to Universiti Teknologi MARA Student Paper | 1% |
| 3 | mafiadoc.com Internet Source | 1% |
| 4 | Masahiro Seino, Ryo Onuma, Hiroki Nakayama, Youzou Miyadera, Shoichi Nakamura. "A framework for supporting the recall of the scenes in complicated documenting processes using screenshots", 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), 2017 Publication | 1% |
| 5 | andryprabowo19.blogspot.com Internet Source | 1% |
| 6 | openaccess.hacettepe.edu.tr:8080 Internet Source | 1% |
| 7 | لطفي ، منيرة محمد مظهر. "تأثير ما وراء البيانات على كفاءة | |

"استرجاع مقالات الدوريات الإلكترونية العربية", Dar al Fajr for
Publishing and Distribution, 2018

Publication

1%

8

iconcs.org

Internet Source

1%

9

journal.uad.ac.id

Internet Source

<1%

10

ascelibrary.org

Internet Source

<1%

11

Peng Zhou, Nora El-Gohary. "Ontology-Based
Multilabel Text Classification of Construction
Regulatory Documents", Journal of Computing
in Civil Engineering, 2016

Publication

<1%

12

digitalia.sbn.it

Internet Source

<1%

13

hдарwanto.blogspot.com

Internet Source

<1%

14

docplayer.info

Internet Source

<1%

15

dzakyghufron.blogspot.com

Internet Source

<1%

16

www.menarik.info

Internet Source

<1%

17

teguhbastian12345.blogspot.com

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography On