

Fuzzy Formal Concept Analysis untuk Kemiripan Dokumen

Endang Supriyati

Progdi Teknik Informatika - UMK

esupriyati@gmail.com

ABSTRAK

Fuzzy logic dapat dimasukkan ke dalam ontologi untuk representasi ketidakpastian informasi yang ditemukan di banyak aplikasi domain karena kurangnya jelas batas-batas antara konsep domain. Fuzzy ontologi dihasilkan dari konsep hirarki yang telah ditetapkan. Namun, untuk membangun sebuah konsep hirarki untuk domain tertentu dapat menjadi tugas yang sulit dan membosankan. Untuk mengatasi masalah ini, diusulkan Fuzzy Formal Concept Analysis (FFCA). Titik awal dari metode diusulkan dalam paper ini adalah definisi dari konteks, relasi kemiripan pada domain ontologi kemudian memetakan ke dalam concept lattice. Dengan penggunaan tool lattice navigator, metode yang diusulkan mampu mengelompokkan domain ontology secara efektif.

Kata Kunci:

Ontology, Formal Concept Analysis, Fuzzy Formal Concept Analysis, konsep Lattice

ABSTRACT

Fuzzy logic can be incorporated into ontology. Fuzzy ontology generated from the concept hierarchy has been determined. However, to construct a concept hierarchy can be a difficult and tedious task. To overcome this problem, this paper use Fuzzy Formal Concept Analysis (FFCA). The starting point is definition of context, similarity relation on the domain ontology and then mapped into the concept lattice. Using of lattice navigator tool, the proposed method can effectively classify ontology domain.

Author Keywords

Ontology, Formal Concept Analysis, Fuzzy Formal Concept Analysis, Lattice concept

Author Keywords

Ontology, Formal Concept Analysis, Fuzzy Formal Concept Analysis, Lattice concept

1 LATAR BELAKANG

Tujuan dari sistem Information Retrieval adalah memenuhi kebutuhan informasi pengguna dengan me-retrieve semua dokumen yang mungkin relevan, pada waktu yang sama me-retrieve sesedikit mungkin dokumen yang tak-relevan. Sistem ini menggunakan fungsi heuristik untuk mendapatkan dokumen-dokumen yang relevan dengan query pengguna. Agar representasi dokumen baik, dokumen-dokumen dengan topik atau isi yang mirip dikelompokkan bersama-sama [14].

Ontology adalah konseptualisasi dari sebuah domain ke dalam pemahaman manusia, dan dapat dibaca oleh mesin dengan format yang berisi entity, atribut, relasi dan aksioma [13]

Fuzzy logic dapat dimasukkan ke ontologi untuk representasi ketidakpastian informasi yang ditemukan di banyak aplikasi domain karena kurangnya jelas batas-batas antara konsep domain.

Secara umum fuzzy ontologi dihasilkan dari konsep hirarki yang telah ditetapkan. Namun, untuk membangun sebuah konsep hirarki untuk domain tertentu dapat menjadi tugas yang sulit dan membosankan. Untuk mengatasi masalah ini, diusulkan Fuzzy Formal Concept Analysis (FFCA) yang menggabungkan logika fuzzy ke dalam Formal Concept Analysis.

Formal Concept Analysis (FCA) [10] adalah teknik analisis data berbasis teori lattice. FCA merepresentasikan relasi antara object dan atribut dalam suatu domain kemudian menginterpretasikannya dalam konsep lattice. Konsep lattice lebih informatif daripada konsep tree tradisional yang juga didukung multiple inheritance.

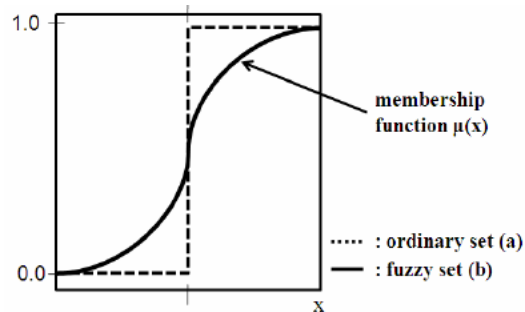
Struktur organisasi dari makalah ini berisi bagian pertama latar belakang, bagian kedua metode yang diusulkan, bagian ketiga uji coba, bagian keempat kesimpulan.

2 METODE YANG DIUSULKAN

Didalam paper ini kami akan mengusulkan sebuah pendekatan FCA dan FFCA untuk mengukur kedekatan antar dokumen.

2.1 Logika Fuzzy

Logika fuzzy, gambar 1, mengizinkan ketidakpastian informasi yang akan diwakili oleh keanggotaan nilai dalam suatu kisaran nilai $(0,1)$ [9]



Gambar 1 : Konsep himpunan Fuzzy

2.2 FCA (Formal Concept Analysis)

Sebuah konteks formal $K = (G, M, I)$ terdiri dua himpunan terhingga obyek G dan himpunan atribut M , dan hubungan biner I antara obyek dan atribut. [10].

FCA memiliki ekstensi dan intensi yang subset objek dan atribut. Ekstensi dan intensi yang diturunkan oleh dua fungsi, yang didefinisikan sebagai:

Definisi 1.

Formal Context (G, M, I) :

G (formal) objects ("things")

M (formal) attributes

$I \subseteq G \times M$ Incidence relation

$intent(A) = \{m \in M \mid \forall g \in A: (g, m) \in I\}$ for $A \subseteq G$,

$extent(B) = \{g \in G \mid \forall m \in B: (g, m) \in I\}$ for $B \subseteq M$.

Sebagai contoh, intensi dari {D1, D3} adalah {a} dan perpanjangan {a} adalah {D1, D3}, karena itu, ({D1, D3}, {a}) adalah sebuah konsep formal. Kita dapat mengekstrak setiap konsep formal dalam Tabel 1, dibawah ini :

Tabel 1. Formal Context

	a	b	c	d
O1	X			X
O2		X	X	
O3	X		X	
O4			X	

Definisi 2

Jika (A_1, B_1) dan (A_2, B_2) adalah dua formal concept dari formal context (G, M, I) . (A_1, B_1) , (A_2, B_2) adalah orde parsial dari super-sub relasi yang dinormalisasi dengan

$$(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subseteq A_2 \iff B_1 \supseteq B_2$$

Definisi 3

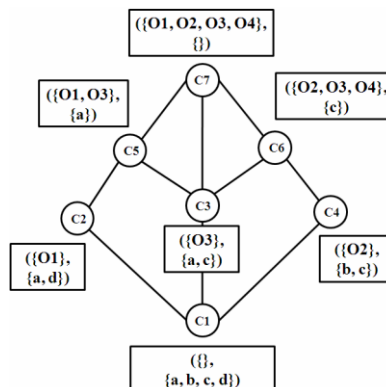
Konsep lattice dari formal context K adalah set $B(C)$ dari seluruh formal concept K dengan order parsial \leq , dinotasikan sebagai :

$$L := (B(C), \leq)$$

Tabel 2. Formal concept dari tabel 1

	Extensions	Intensions
C1	{}	{a, b, c, d}
C2	{O1}	{a, d}
C3	{O3}	{a, c}
C4	{O2}	{b, c}
C5	{O1, O3}	{a}
C6	{O2, O3, O4}	{c}
C7	{O1, O2, O3, O4}	{}

Tabel 2, merupakan formal concept dari tabel 1, gambar 2 merupakan konsep latiiice dari tabel1.



Gambar2: Konsep Lattice dari Tabel 1

2.3 FFCA(Fuzzy Formal concept Analysis)

Seperti yang didefinisikan [11], $K:=(G,M,I=\psi(G \times M))$ dimana G adalah himpunan terhingga objek, M adalah finite set of atribut, dan I adalah himpunan fuzzy di domain $G \times M$. Setiap relasi $(g, m) \in I$ mempunyai sebuah nilai anggota $\mu(g, m)$ dalam $[0,1]$.

Definisi 4

Untuk fuzzy formal context $K:=(G,M,I)$ dan sebuah cofidence threshold $T=[t_1,t_2]$, didefinisikan $FI(A)=\{m \in M \mid \forall g \in A: t_1 \leq \mu(g,m) \leq t_2\}$ untuk $A \subseteq G$ dan $FE(B)=\{g \in G \mid \forall m \in B: t_1 \leq \mu(g,m) \leq t_2\}$ untuk $B \subseteq M$.

Tabel 3. Fuzzy Formal Context

	IR(a)	DM(b)	DB(c)	ML(d)
O1	0.8	0.12	0.61	0.6
O2	0.9	0.85	0.13	0.1
O3	0.1	0.14	0.87	0.1
O4	0.6	0.12	0.13	0.3

Tabel 3, merupakan Fuzzy Formal Context yang merupakan hasil dari nilai Fuzzy yang dimasukkan ke dalam formal concept.

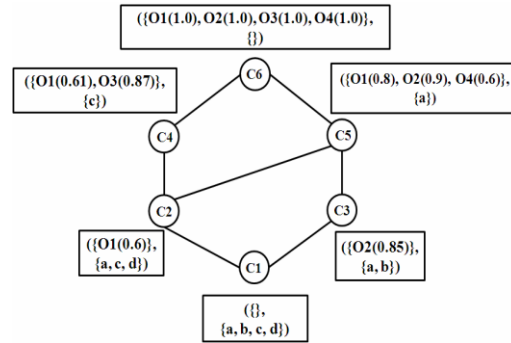
Tabel 4. Fuzzy Formal Context dg $T = [0.5-1.0]$

	IR(a)	DM(b)	DB(c)	ML(d)
O1	0.8	-	0.61	0.6
O2	0.9	0.85	-	-
O3	-	-	0.87	-
O4	0.6	-	-	-

Tabel 4, merupakan Fuzzy formal concept dengan nilai Fuzzy antara 0.5 sampai 1, tabel 5 adalah hasil Fuzzy formal context dari tabel 4. Gambar 3 merupakan konsep lattice dari tabel 5.

Tabel 5. Fuzzy Formal Context Tabel 4

	Extens dg nilai anggota	Intensions
C1	{}	{a, b, c, d}
C2	{O1(0.6)}	{a,c, d}
C3	{O2(0.85)}	{a, b}
C4	{O1(0.61),O3(0.87)}	{c}
C5	{O1(0.8),O2(0.9), O4(0.6)}	{a}
C6	{O1(1.0),O2(1.0),O3(1.0),O4(1.0)}	{}



Gambar3: Konsep Lattice dari Tabel 5

2.4 Fuzzy Ontology

Logika fuzzy dapat dimasukkan ke dalam ontology untuk merepresentasikan ketidakpastian informasi. [3] biasanya *fuzzy ontology* dihasilkan dari predefinisi konsep hirarki. *Fuzzy ontology* F_o terdiri dari empat elemen (C, A^c, R, X), dimana C merepresentasikan konsep, A^c merepresentasikan koleksi atribut, satu untuk setiap konsep, dan $R = (R_T, R_N)$ merepresentasikan relasi, yang terdiri dari dua elemen: R_N adalah sepasang relasi nontaksonomi dan R_T adalah relasi taksonomi. X adalah aksioma

$C = \{\text{"Document"}, \text{"Research Area"}\}$

$AC(\text{"Document"}) = \{\text{"Name"}, \text{"Author"}, \text{"Title"}, \text{"Keywords"}, \text{"Abstract"}, \text{"Body"}, \text{"Publisher"}, \text{"Publication Date"}\}$

$AC(\text{"Research Area"}) = \{\text{"Name"}, \text{"Keyword"}\}$

$RN = \{\text{belong-to}(\text{"Document"}, \text{"Research Area"}), \text{consist-of}(\text{"Research Area"}, \text{"Document"})\}$

$RT = \{\text{superarea-of}(\text{"Research Area"}, \text{"Research Area"}), \text{subarea of}(\text{"Research Area"}, \text{"Research Area"})\}$

$X = \{\text{Implies}(\text{Antecedent}(\text{consist-of}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{belong-to}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

$\text{Implies}(\text{Antecedent}(\text{belong-to}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{consist-of}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

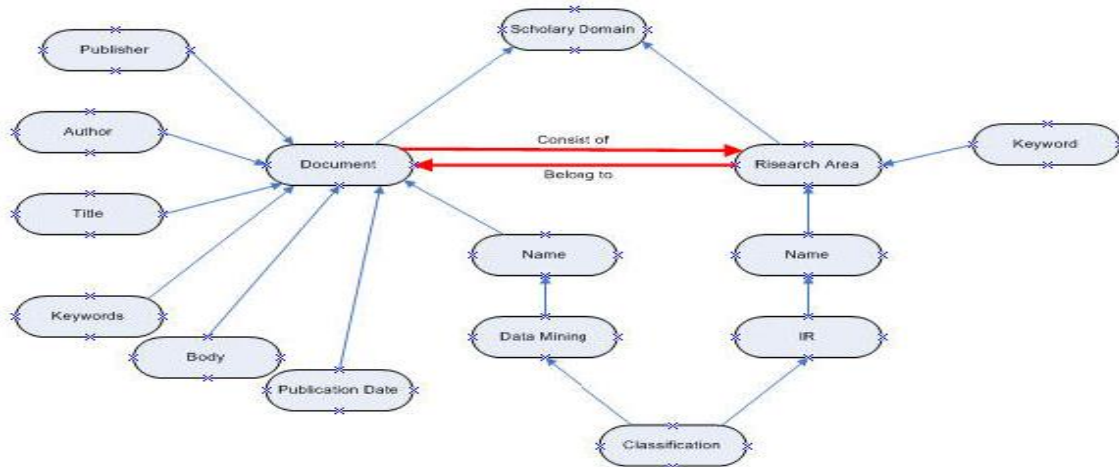
$\text{Implies}(\text{Antecedent}(\text{superarea}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{subarea}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

$\text{Implies}(\text{Antecedent}(\text{subarea}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{superarea}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

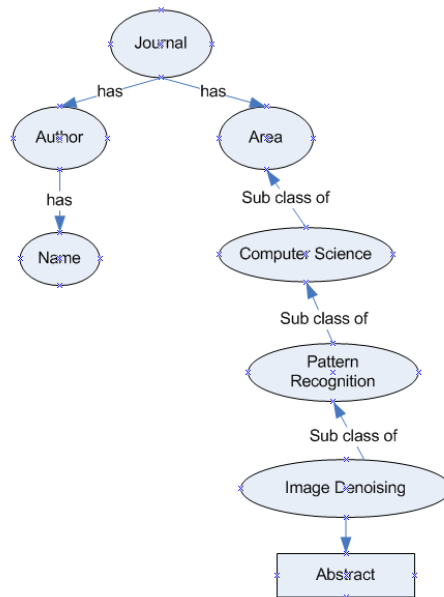
$\text{Implies}(\text{Antecedent}(\text{subarea}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{superarea}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

$\text{Implies}(\text{Antecedent}(\text{superarea}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{subarea}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$

$\text{Implies}(\text{Antecedent}(\text{superarea}(\text{I-variable}(x1) \text{I-variable}(x2))) \text{Consequent}(\text{subarea}(\text{I-variable}(x2) \text{I-variable}(x1))))\}$



Gambar 4 : Non-Taxonomy Ontology



Gambar 5. Taxonomy Ontology

Gambar 4 dan 5 merepresentasikan relasi nontaksonomi dan relasi taksonomi dari scholarly domain.

3 UJI COBA

Data yang digunakan adalah data scholarly yang didownload dari[8]. Tool yang digunakan diperoleh dari [15] yang merupakan situs lattice navigator.

Kemudian dipilih term untuk membangun domain ontologi dengan menghitung bobot term dengan TF-IDF[12].

Bobot lokal suatu term i di dalam dokumen j (tf_{ij}) dapat didefinisikan sebagai:

$$tf_{ij} = \frac{f_{ij}}{\max_i(f_{ij})}$$

Dimana f_{ij} adalah jumlah berapa kali term i muncul di dalam dokumen j .

Bobot global dari suatu term i pada pendekatan inverse document frequency (idf_i) dapat didefinisikan sebagai :

$$idf_i = \log \left(\frac{n}{df_i} \right)$$

Dimana df_i adalah frekuensi dokumen dari term i dan sama dengan jumlah dokumen yang mengandung term i .

Dari total dokumen yang ada, hanya diambil beberapa saja, untuk menunjukkan hasil dari metode yang di usulkan secara jelas.

Tabel 6. Frekuensi Term

	Software	opensource	Networking	web
D1	30	23	5	9
D2	4	7	15	50
D3	11	4	40	11
D4	10	36	15	24

Tabel 6 merupakan tabel frekuensi term, tabel ini diperoleh dari perhitungan dari perhitungan manual.

Tabel 7. Total Frekuensi

	Total Term
Software	223
opensource	230
Networking	185
web	218

Tabel 7 menunjukkan total frekuensi yang di hitung, tabel 8 menunjukkan bobot term pada domain yang di hitung berdasarkan TF-IDF.

Tabel 8. Bobot term pada domain

	Software	opensource	Networking	web
D1	0,56	0,41	0,12	0,17
D2	0,07	0,13	0,36	0,96
D3	0,21	0,07	0,96	0,21
D4	0,19	0,64	0,36	0,46

Tabel 9. Fuzzy Threshold

T1	0 - 0.4	rendah
T2	0.41 - 0.79	sedang
T3	0.80 - 1.00	tinggi

Pada makalah ini, di tentukan nilai threshold [0,40] sebagai kategori rendah, [0,41-0,79] sebagai kategori sedang (medium) dan nilai [0,80-1] sebagai kategori tinggi seperti yang terlihat pada tabel 9.

Tabel 10. Fuzzy context, $T=[0 - 0.40]$

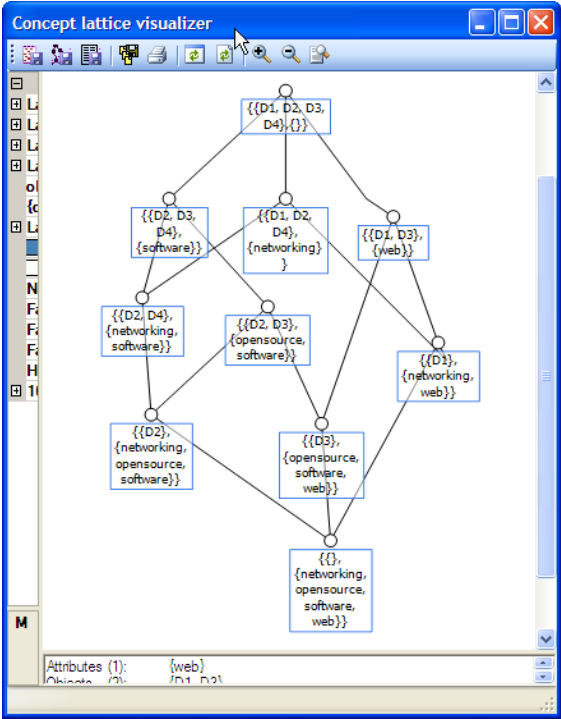
Obj/Attr	software	opensource	networking	web
D1	-	-	0,12	0,17
D2	0,07	0,13	0,36	-
D3	0,21	0,07	-	0,21
D4	0,19	-	0,36	-

Tabel 10 menunjukkan hasil fuzzy context untuk threshold kategori rendah yaitu 0 -0.40. Dengan tool diperoleh kedekatan atau kemiripan dokumen secara extent dan intent lattice concept seperti yang terlihat pada gambar 6.

ConceptID	Extent	Intent
c(0)	{D1; D2; D3; D4}	{}
c(1)	{D1; D3}	{Web}
c(2)	{D1; D2; D4}	{Networking}
c(3)	{D1}	{Networking; Web}
c(4)	{D2; D3; D4}	{Software}
c(5)	{D2; D4}	{Software; Networking}
c(6)	{D2; D3}	{Software; OpenSource}
c(7)	{D3}	{Software; OpenSource; Web}
c(8)	{D2}	{Software; OpenSource; Networking}
c(9)	{}	{Software; OpenSource; Networking; Web}

Gambar 6. Extent dan Intent Lattice Concept

Gambar 7, menunjukkan visualisasi fuzzy formal concept yang berbentuk diagram lattice.



Gambar 7. Screenshot Fuzzy Format Concept Analysis

4. KESIMPULAN

Pada makalah ini, diusulkan Fuzzy Formal Concept Analysis sebagai pendekatan pengelompokan dokumen secara konseptual. Pendekatan yang diusulkan terdiri dari langkah-langkah berikut : Formal Concept Analysis, Fuzzy Formal Concept Analysis, pengelompokan dokumen berdasar threshold tertentu. Dari metode *Formal Concept Analysis* dan *Fuzzy Formal Concept Analysis* dihasilkan data yang menginformasikan kemiripan antar dokumen. Dengan penggunaan tool lattice navigator metode yang diusulkan mampu mengelompokkan domain ontology secara efektif.

REFERENCES

- [1] D. Fensel, F. van Harmelen, I. Horrocks, D.L. McGuinness, and P.F. Patel-Schneider, "OIL: an Ontology Infrastructure for the Semantic web.," IEEE Intelligent Systems, vol. 16, no. 2, pp. 38-45, 2001.
- [2] W3C, Resource description framework (RDF): concepts and abstract syntax, in: G. Klyne, J.J. Carroll, B. McBride (Eds.), W3C Recommendation, 10, February 2004. <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>
- [3] Quan Thanh Tho, Siu Cheung Hui, "Automatic Fuzzy Ontology Generation for Semantic Web", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 6, JUNE 2006
- [4] Silvia Calegari and Davide Ciucci, "INTEGRATING FUZZY LOGIC IN ONTOLOGIES", Dipartimento di Informatica, Sistemistica e Comunicazione.
- [5] Fereshteh Mahdavi, Maizatul Akmar Ismail, and Noorhidawati Abdullah, "Semi-Automatic Trend Detection in Scholarly Repository Using Semantic Approach", World Academy of science Engineering and Technology 52 2009.
- [6] ANDREWS, Simon J, Data conversion and interoperability for FCA Available from Sheffield Hallam University Research Archive (SHURA) at: <http://shura.shu.ac.uk/36/>
- [7] Ben Martin, Formal Concept Analysis and Semantic File System, University of Wolonggong, 2008
- [8] ww.csd.abdn.ac.id/~ggrimnes/download/citeseer_dadta.tar.gz
- [9] L.A Zadeh. Fuzzy Sets. Journal of Information and Control, Vol. 8. 1965, pp.338-353.
- [10] B. Ganter and R. Wille, Formal Concept Analysis Mathematical foundations, Springer-Verlag Berlin Heidelberg, 1999
- [11] T. T. Quan, S. C. Hui and T. H. Cao, "A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation
- [12] Salton, G.. Automatic text processing. Chapter 9, 1989

- [13] N. Guarino and P. Giaretta, *Ontologies and Knowledge Bases: Towards a Terminological Clarification. Toward Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. Amsterdam: IOS Press, 1995.
- [14] Murad, Azmi MA., Martin, Trevor. (2007) *Word Similarity for Document Gouping using Soft Computing*. IJCSNS International Journal of Computer Science and Network Security, Vol.7 No.8, August 2007, pp. 20- 27
- [15] <http://www.fca.radvansky.net/>

BIODATA DIRI

Nama : Endang Supriyati
NIDN : 0629077402
Jurusan : Teknik Informatika
Fakultas : Teknik
Universitas : Universitas Muria Kudus (UMK)
Email : esupriyati@gmail.com
Publikasi Sebelumnya : -