

RANCANG BANGUN PENCARIAN JUDUL TESIS BERBASIS TEKNOLOGI WEB SEMANTIK

Ahmad Chusyairi¹⁾, Ema Utami²⁾

^{1,2)} Magister Teknik Informatika Program Pascasarjana STMIK AMIKOM Yogyakarta
Jl Ring Road Utara, Condongcatur, Sleman, Yogyakarta 55281
Email : ahmad.c@students.amikom.ac.id¹⁾, ema.u@amikom.ac.id²⁾

Kecepatan akses informasi telah menjadi suatu kebutuhan khususnya bagi dunia pendidikan. Bagian front office adalah pengguna yang tidak mempunyai latar belakang pengetahuan query yang cukup, sehingga kebutuhan pengguna adalah kemudahan untuk memperoleh informasi terkait dengan pencarian judul tesis dengan menggunakan bahasa alami yaitu bahasa Indonesia. Rancang bangun pencarian dengan sistem pencarian thesaurus dengan pendekatan SKOS (Simple Knowledge Organization System) yang merupakan bagian dari teknologi web semantik dan pendekatan NLP (Natural Language Processing). Hasil yang diharapkan adalah mampu memberikan informasi kepada pengguna, sehingga penulisan judul tesis sesuai dengan aturan yang ada.

Kata kunci: Pencarian judul tesis, thesaurus, SKOS, web semantik, NLP

1. Pendahuluan

Informasi judul tesis yang ditampilkan pada dokumen web (*web content*) Magister Teknik Informatika STMIK AMIKOM Yogyakarta hanya dapat dipahami oleh manusia namun tidak dapat dipahami oleh mesin, sehingga mesin tidak mampu menginterpretasikan informasi apa yang dibutuhkan atau dicari oleh manusia. Hal ini mengakibatkan dokumen-dokumen yang tidak relevan disertakan sebagai hasil pencarian (*search result*). Dan sering terjadi dokumen-dokumen yang relevan justru tidak terindeks oleh mesin pencari, sehingga campur tangan manusia untuk memilah informasi-informasi tersebut tetap dibutuhkan. Untuk mengatasi kesulitan tersebut, dibutuhkan suatu cara agar informasi dalam suatu dokumen web dapat dibaca dan dipahami oleh mesin (*machine understandable*). Web yang memiliki kecerdasan dalam memberikan jawaban atas pertanyaan tentang judul tesis para penggunanya. [1]

Semantic web (SW) yang dipelopori oleh Tim Berners-Lee, merupakan suatu cara untuk merepresentasikan *web content* dalam bentuk yang dapat dipahami dan diproses oleh mesin. Dengan kata lain, SW mengidentifikasi bahwa makna data (*the meaning of data*) pada web dapat dipahami, baik oleh manusia maupun oleh mesin (komputer).

Dalam pencarian informasi pada sebuah basis data, selain diperlukan pemahaman dalam bahasa Inggris, juga diperlukan pemahaman yang memadai tentang struktur penulisan query dan pemahaman antar muka suatu sistem basis data. Untuk itu diperlukan suatu sistem yang dapat digunakan untuk melakukan pencarian data dengan cara yang lebih sederhana, daripada menuliskan pernyataan query yang cenderung tidak dipahami oleh pengguna awam. Sistem yang dibutuhkan adalah yang dapat memberikan jawaban atas pertanyaan yang diajukan oleh pengguna awam dengan memasukkan pertanyaan dalam bahasa alami yang dikuasai oleh pengguna. Sistem ini juga harus mempunyai tampilan antarmuka yang bersifat *user friendly*, sehingga mudah dipahami dan digunakan oleh pengguna awam. [2]

Rancang bangun pencarian *thesaurus* dengan pendekatan SKOS (*System Knowledge Organization System*) yang merupakan bagian dari teknologi web semantik. SKOS efektif digunakan sebagai salah satu pendekatan berbasis kosakata untuk padanan kata (*thesauri*), klasifikasi dan daftar istilah (*glossaries*), yang dimungkinkan sebuah dokumen dapat dihubungkan dengan dokumen yang lainnya, sehingga dokumen tersebut tidak lagi berdiri sendiri melainkan sebuah dokumen yang saling memiliki relasi. SKOS juga digunakan untuk merepresentasikan semantik antar sumber daya sedangkan pendekatan NLP (*Natural Language Processing*) digunakan untuk mengolah kalimat sebagai kata kunci pencarian. Dari beberapa perpaduan teknologi web semantik dan NLP diatas mampu memberikan hasil kepada pengguna dengan informasi yang disajikan lebih kaya. [3]

Adapun rumusan masalah dalam penelitian ini adalah:

- Bagaimana membuat aturan bahasa untuk rancang bangun pemrosesan bahasa alami, yaitu bahasa Indonesia dengan SKOS (*Simple Knowledge Organization Sytem*) dan NLP (*Natural Language Processing*) pada basis data tesis.
- Bagaimana membuat prototipe perangkat lunak rancang bangun pemrosesan bahasa alami, yaitu bahasa Indonesia.

Tujuan penelitian ini adalah membuat rancang bangun untuk pencarian judul tesis berbasis web semantik dengan menggunakan bahasa alami, yaitu bahasa Indonesia.

Persamaan dari penelitian sebelumnya adalah menggunakan teknologi *web semantik* untuk membangun sistem rancang bangun pencarian sedangkan perbedaannya hanya terletak pada subyek (*domain*) dan metode sistem pencariannya saja, yang dalam hal ini menggunakan *thesaurus* dari beberapa penelitian yang masih menggunakan sistem berbasis pola.

Metode yang akan digunakan dalam penelitian ini terdiri dari langkah-langkah sebagai berikut:

a. Obyek penelitian

Obyek penelitian dari penelitian ini adalah *web content* Magister Teknik Informatika STMIK AMIKOM Yogyakarta.

b. Data yang diperlukan

Merupakan data yang mendukung dalam penelitian ini meliputi data primer dan data sekunder.

1) Data primer

Data berupa basis data tesis khususnya judul tesis pada semua konsentrasi (CIO Management, Sistem Informasi, dan Teknologi Media Digital) di Magister Teknik Informatika STMIK AMIKOM Yogyakarta.

2) Data sekunder

Data yang diperoleh dengan membaca dan mempelajari referensi mengenai pengolahan bahasa alami, teknologi web semantik, bahasa pemrograman Java serta perancangan query dengan menggunakan SPARQL.

c. Teknik pengumpulan data

Tahap ini mempunyai tujuan untuk mendapatkan data-data tesis yang mempunyai keterkaitan dengan topik penelitian, sehingga lebih relevan dan akurat. Teknik pengumpulan data yang dilakukan pada penelitian ini adalah sebagai berikut:

1) Observasi

Dengan melakukan pengamatan dan pencatatan secara sistematis tentang hal-hal yang berhubungan dengan judul tesis.

2) Studi pustaka

Teknik pengumpulan data dilakukan dengan mengumpulkan bahan-bahan, referensi, dan dokumen yang berhubungan dengan judul tesis.

3) Metode pengembangan

Penelitian ini menggunakan metode *prototyping*, dalam model ini sistem pencarian judul tesis dirancang dan dibangun secara bertahap, pada setiap tahap pengembangan dilakukan percobaan untuk melihat apakah sistem sudah bekerja sesuai dengan yang diinginkan.

Berikut adalah tahapan yang dilakukan pada penelitian ini dengan menggunakan metode pengembangan *prototyping*:

a) Analisis

Pada tahap ini dilakukan analisis tentang masalah penelitian yang terkait dengan judul tesis dan menentukan pemecahan masalah yang tepat untuk menyelesaikannya.

b) Desain

Pada tahap ini dibangun rancangan proses dari sistem dengan menggunakan diagram bantu seperti UML dan *tools* Protégé 4.0.2.

c) Prototipe

Pada tahap ini dibangun rancang bangun *semantic web* dengan menggunakan pendekatan pemrosesan bahasa alami dalam melakukan pencarian judul tesis. Desain *content* dan manajemen menu menggunakan *Java Server Page* (JSP) dan bahasa pemrograman Java.

d) Pengujian

Pada tahap ini dilakukan pengujian pada rancang bangun pencarian judul tesis. Pengujian dilakukan dengan validasi aturan kalimat masukan bagi sistem dan keluaran yang diharapkan.

e) Evaluasi

Pada tahap ini dilakukan evaluasi apakah performa rancang bangun pencarian judul tesis sudah sesuai dengan yang diharapkan, apabila belum sesuai akan dilakukan penyesuaian secukupnya.

f) Penyesuaian

Tahap ini dilakukan apabila pada tahap evaluasi, performa rancang bangun kurang memadai dan dibutuhkan perbaikan. Tahap ini melakukan penyesuaian dan perbaikan pada rancang bangun pencarian judul tesis sesuai dengan kebutuhan. [3]

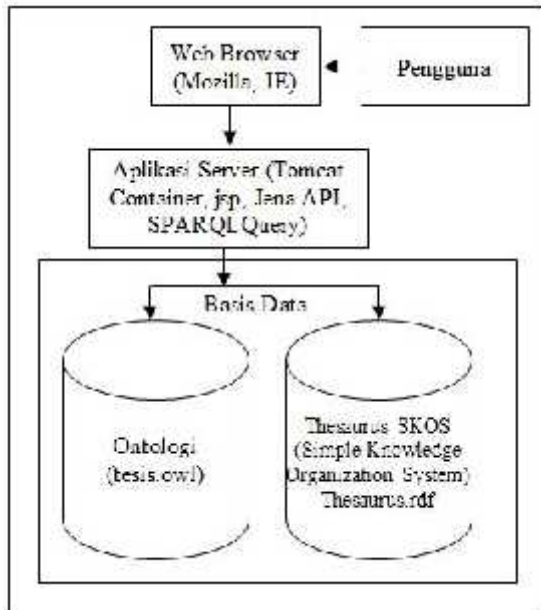
2. Pembahasan

Penelitian ini bertujuan untuk membangun dan merancang sebuah prototipe rancang bangun pencarian judul tesis berbasis web semantik. Rancang bangun sistem ini memberikan beberapa fasilitas, yaitu:

- Menyediakan fasilitas pencarian judul tesis dari semua konsentrasi di Magister Teknik Informatika STMIK AMIKOM Yogyakarta berupa layanan pencarian berdasarkan kategori, berupa judul, penulis dan tahun terbit tesis dengan menggunakan kunci pencarian berdasar judul dan penulis dari koleksi yang tersedia.
- Menyediakan fasilitas pencarian judul tesis dalam bentuk layanan pencarian dengan menggunakan bahasa alami berupa kalimat perintah pencarian berbahasa Indonesia.

- c. Menyediakan fasilitas untuk menampilkan keseluruhan koleksi tesis dari semua konsentrasi berdasarkan kategori judul dan penulis.

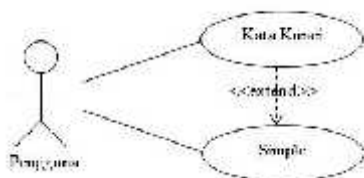
Berikut adalah gambaran umum Rancang Bangun Pencarian Judul Tesis Berbasis Web Semantik seperti pada gambar 1. Pengguna membuka *Web Browser* (Mozilla Firefox, Internet Explore, dll.), maka akan tampil aplikasi server seperti Tomcat Container, jsp, Jena API, SPARQL Query. Aplikasi server tersebut akan menampilkan informasi dari database ontologi (tesis.owl) dan *thesaurus* SKOS (thesaurus.rdf). [4]



Gambar 1. Gambaran Umum Rancang Bangun Pencarian Judul Tesis Berbasis Teknologi Semantik Web

Pengguna dimungkinkan dapat memilih beberapa menu pencarian yang ada, seperti pencarian kata kunci dan simple. Notasi “extend” memberikan alternatif apabila dengan *keyword searching* tidak memberikan hasil, sehingga pengguna dapat memilih menu yang lain.

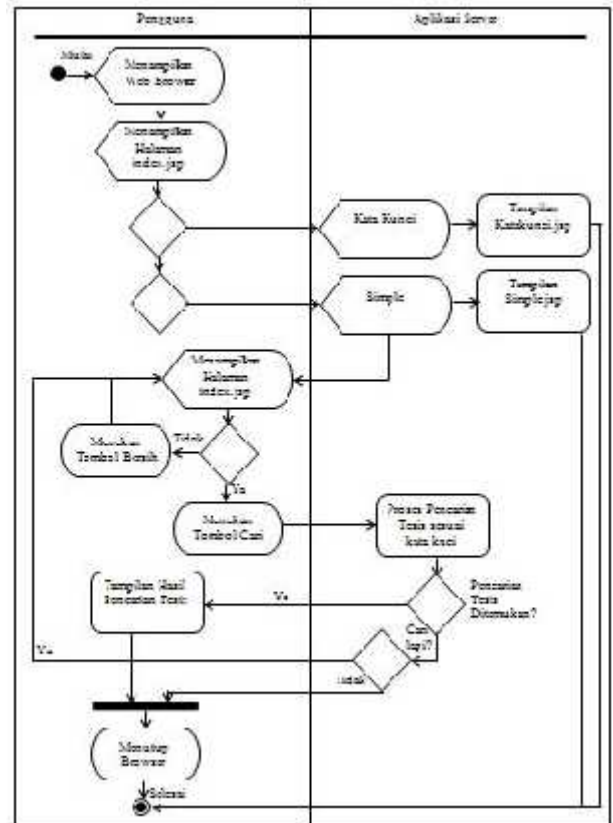
Diagram Use Case untuk Pencarian Judul Tesis dijelaskan pengguna dapat mencari berdasarkan kata kunci dan simple pada Gambar 2. [4]



Gambar 2. Diagram Use Case Pencarian Judul Tesis

Diagram Aktifitasnya untuk Pencarian Judul Tesis dijelaskan mulai pengguna membuka *Web Browser* lalu menampilkan halaman *index.jsp* dan mencari judul tesis berdasarkan kata kunci dan simple, sehingga dapat menampilkan *katakunci.jsp* dan *simple.jsp* pada server.

Untuk pencarian simple pengguna memasukkan kata kunci jika pilihannya “tidak” dengan menekan tombol bersih dan jika pilihannya “ya” dengan meneka tombol cari, maka oleh server akan memproses pencarian tesis sesuai kata kunci jika pencariannya ditemukan “ya”, maka akan ditampilkan hasil pencarian tesisnya dan jika “tidak” apakah akan mencari lagi? Jika “ya” masukan kata kuncinya dan jika “tidak” dengan menutup web browsernya seperti pada Gambar 3. [4]



Gambar 3. Diagram Aktifitas Pencarian Tesis

Adapun perancangan dari Rancang Bangun Pencarian Judul Tesis Berbasis Web Semantik terdapat dua cara, yaitu menggunakan Kata Kunci, yaitu pencarian berdasarkan Judul seperti pada Gambar 4. [4]



Gambar 4. Rancangan Antarmuka Pencarian Berdasar Kata Kunci Rancang Bangun Judul Tesis

Dan cara kedua Rancang Bangun Pencarian Judul Tesis Berbasis Web Semantik adalah Simple, yaitu pencarian berdasarkan Judul dan Penulis seperti pada Gambar 5. [4]



Gambar 5. Rancangan Antarmuka Pencarian Berdasar Simple Rancang Bangun Tesis

Aturan produksi yang dibuat adalah aturan produksi yang dapat mengakses data karena pola aturan produksi yang sudah ditentukan adalah pola aturan produksi yang sesuai dengan pola *query* SPARQL untuk mengakses basis pengetahuan ontologi judul tesis. Perancangan aturan produksi dapat dilakukan dengan menentukan lebih dahulu pola keteraturan perintah untuk mengakses ontologi judul tesis. Pola keteraturan aturan produksi mengandung lima unsur, yaitu:

- Kata perintah pencarian,
- Kata kunci yang mewakili kategori objek pencarian seperti judul, penulis,
- Kata kunci yang mewakili *datatype property*,
- Kata kunci yang mewakili *object property*, dan
- Nilai dari sebuah *datatype property*. [3]

Sesuai dengan kelima unsur yang ada, dapat dibentuk tipe kalimat pencarian yang dapat diproses oleh sistem. Penentuan tipe kalimat pencarian dalam memberikan masukan bagi sistem akan sangat menentukan dalam penggolongan tipe *query* SPARQL bagi sistem. Simbol-simbol yang digunakan dalam penulisan tipe kalimat masukan bagi sistem adalah sebagai berikut:

- [kp], kata perintah yang mewakili kalimat masukan.
- [ctg], simbol yang digunakan untuk mewakili kategori objek yang dicari.
- [dtp], simbol yang mewakili *datatype property*.
- [obp], simbol yang mewakili *object property*.
- [value], simbol yang digunakan untuk mewakili nilai dari sebuah *datatype property* atau *object property*.

Berdasarkan bentuk pertanyaan kalimat pencarian yang dijadikan masukan bagi sistem untuk melakukan pencarian, dapat diidentifikasi dalam bentuk pola kalimat (K), sebagai berikut:

$K \rightarrow kp + ctg + [dtp|obp] + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$

$dtp \rightarrow \{judul\}$
 $obp \rightarrow \{penulis\}$
 $value \rightarrow \{?\}$

Dari pola diatas akan terbentuk 6 buah tipe kalimat pencarian, yaitu:

- Tipe I : $K \rightarrow kp + ctg + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $value \rightarrow \{?\}$
- Tipe II : $K \rightarrow kp + ctg +.dtp + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $dtp \rightarrow \{judul\}$
 $value \rightarrow \{?\}$
- Tipe III : $K \rightarrow kp + ctg + obp + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $obp \rightarrow \{penulis\}$
 $value \rightarrow \{?\}$
- Tipe IV : $K \rightarrow kp + ctg +.dtp + value + obp + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $dtp \rightarrow \{judul\}$
 $value \rightarrow \{?\}$
 $obp \rightarrow \{penulis\}$
 $value \rightarrow \{?\}$
- Tipe V : $K \rightarrow kp + ctg +.dtp + value +.dtp + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $dtp \rightarrow \{judul\}$
 $value \rightarrow \{?\}$
 $dtp \rightarrow \{judul\}$
 $value \rightarrow \{?\}$
- Tipe VI : $K \rightarrow kp + ctg + obp + value + obp + value$
 $kp \rightarrow \{cari|tampilkan\}$
 $ctg \rightarrow \{tesis\}$
 $obp \rightarrow \{penulis\}$
 $value \rightarrow \{?\}$
 $obp \rightarrow \{penulis\}$
 $value \rightarrow \{?\}$

Pengujian terhadap sistem rancang bangun pencarian judul tesis berbasis web semantik untuk mengetahui tingkat keberhasilan sistem dalam memahami pertanyaan yang disampaikan dengan menggunakan kalimat tanya dalam bahasa Indonesia. Pengujian dilakukan dengan dua cara, yaitu secara kualitatif dan kuantitatif. [5]

Pengujian secara kualitatif dilakukan dengan pertanyaan-pertanyaan yang dipilih untuk mengetahui kemampuan sistem dalam memahami pertanyaan yang disampaikan dengan kalimat yang sesuai dengan kaidah tata bahasa Indonesia. Pengujian untuk mengetahui kemampuan sistem dalam memahami kalimat tunggal yang sesuai dengan kaidah tata bahasa Indonesia. Contoh dari kalimat tanya yang berpola Subjek-Predikat (S-P atau P-S), yang digunakan dalam pengujian ini antara lain: Siapa penulis tesis yang berjudul Pemanfaatan

Artificial Intelligence Markup Language dalam Membangun Chatbot untuk Tanya Jawab Seputar Hukum Faraid ?

Gambar 6 baris 21-40 merupakan log yang dihasilkan ketika sistem mentransformasikan representasi sintaksis kalimat tanya menjadi informasi semantik dalam bentuk query SPARQL. [5]

```

21 JUDULG (SemanticAnalyses.java:96) - transform
22 [[[slapakah]] [[penulis]] [[tesis]] [[Pemanfaatan
23 Artificial Intelligence Markup Language dalam
24 Membangun ChatBot untuk Tanya Jawab Seputar
25 Hukum Faraid]]]] into
26 query SPARQL:
27 PREFIX ds: <https://purl.org/dc/terms/>
28 PREFIX lueli: <https://amikom.stmi.ac.id/>
29 PREFIX opt: <https://www.amikom.ac.id/>
30 SELECT DISTINCT ?doc ?val
31 WHERE {
32 ?val ?doc uruzesaur ?res.
33 ?res URUB a opt:tesas.
34 (?res ds:has distinct ?doc ds:has) IMITON
35 (?res ds:has distinct ?doc ds:has)
36 ?res isothiname ?val.
37 FILTER REGEX(?res ds:fb, "Pemanfaatan Artificial
38 Intelligence Markup Language dalam Membangun
39 ChatBot untuk Tanya Jawab Seputar Hukum Faraid", "i").
40 }
    
```

Gambar 6. Log representasi sintaksis kalimat tanya menjadi informasi semantik dalam bentuk query SPARQL

Pengujian secara kuantitatif dilakukan untuk mengetahui jumlah pertanyaan yang dapat dipahami atau dideteksi sebagai kalimat yang tidak sesuai dengan kaidah tata bahasa Indonesia. [5]

Tabel 1. Contoh pertanyaan yang tidak dapat dipahami oleh sistem beserta penyebabnya

No	Pertanyaan	Penyebab pertanyaan tidak dapat dipahami oleh sistem
1.	Apa saja tesis yang ada?	Kata <i>ada</i> tidak terdapat dalam kosakata sistem
2.	Siapa saja pembimbing yang terlibat dalam pembuatan tesis?	Kata <i>terlibat</i> dan <i>pembuatan</i> tidak terdapat dalam kosakata sistem
3.	Tesis apa yang direferensi Bayu Setiaji ?	Menurut pengetahuan yang dimiliki sistem, orang tidak dapat mengacu pada laporan tesis

3. Kesimpulan

Dari pembahasan diatas dapat disimpulkan:

- a. SKOS (Simple Knowledge Organization System) dan NLP (Natural Language Processing) untuk pengelolaan kosakata, sehingga dapat mengetahui hubungan antar dokumen dalam pencarian judul tesis.
- b. Pendekatan *thesaurus* dengan SKOS dapat mendeteksi pertanyaan yang disampaikan dengan

menggunakan kalimat tanya yang tidak sesuai dengan kaidah tata bahasa Indonesia.

Adapun saran dari penulis adalah:

- a. Penerapan ontologi pada domain selain rancang bangun pencarian judul tesis berbasis web semantik yang merupakan langkah awal dari penerapan web semantik yang masih sederhana, sehingga untuk pengembangan selanjutnya diharapkan yang lebih kompleks dari sisi ontologinya.
- b. Tersedianya data inputan untuk data tesis baru, sehingga tidak manual.

Daftar Pustaka

- [1] Bendi, R.K.J., 2009, Sistem Questio Answering Sederhana Berbasis Ontologi Sebagai Aplikasi Web Semantik, *Tesis*, Program Pascasarjana S2 Ilmu Komputer, Universitas Gadjah Mada Yogyakarta
- [2] Wibisono, S., 2010, Aplikasi Pengolah Bahasa Alami untuk Query Basis Data Akademik dengan Format Data XML, *Tesis*, Program Pascasarjana S2 Ilmu Komputer, Universitas Gadjah Mada Yogyakarta
- [3] Andri, 2011, Rancang Bangun Online Access Catalog (OPAC) Berbasis Web Semantik, *Tesis*, Program Pascasarjana S2 Ilmu Komputer, Universitas Gadjah Mada Yogyakarta
- [4] Nurkhamid, M., 2009, Aplikasi Bibliografi Perpustakaan Berbasis Teknologi Web Semantik, *Tesis*, Program Pascasarjana S2 Ilmu Komputer, Universitas Gadjah Mada Yogyakarta
- [5] Dharma S.I.W., 2013, Sistem Question Answering Menggunakan Pendekatan Berbasis Pengetahuan, *Tesis*, Program Pascasarjana S2 Ilmu Komputer, Universitas Gadjah Mada Yogyakarta

Biodata Penulis

Ahmad Chusyairi, S.Kom, memperoleh gelar Ahli Madya (A.Md) dari Manajemen Informatika STMIK AMIKOM Yogyakarta dengan predikat cumlaude pada tahun 2005. Tahun 2006 memperoleh gelar Sarjana Komputer (S.Kom) dari Program Studi Sistem Informasi STMIK AMIKOM Yogyakarta. Saat ini sebagai mahasiswa (semester 1) Magister Teknik Informatika Program Pasca Sarjana STMIK AMIKOM Yogyakarta. Sejak 2013 menjadi staf akademik dan pengajar di Politeknik PPKP Yogyakarta.

Dr. Ema Utami, S.Si, M.Kom, memperoleh gelar Sarjana Sains (S.Si) dari Program Studi Ilmu Komputer UGM pada tahun 1997. Tahun 2002 memperoleh gelar Magister Komputer (M.Kom) dengan predikat cumlaude dari Program Pascasarjana Ilmu Komputer UGM. Tahun 2010 memperoleh gelar Doktor dari Program Doktor Ilmu Komputer UGM. Sejak 1998 menjadi Staff Pengajar di STMIK AMIKOM Yogyakarta dan sejak 2010 menjadi Wakil Direktur I Bidang Akademik Program Pascasarjana STMIK AMIKOM Yogyakarta.

