PENGGUNAAN *DIGITAL SIGNATURE* PADA METADATA UNTUK PENCARIAN PUBLIKASI ILMIAH BERBASIS SEMANTIK

Nova Hadi Lestriandoko[1)] , Taufiq Wirahman[2)]
Pusat Penelitian Informatika – Lembaga Ilmu Pengetahuan Indonesia
Kompleks LIPI Gedung 20 Lantai 3
Jl. Sangkuriang No.21/154D Bandung 40135
Telepon (022) 2504711 Faks: (022) 2504712
E-mail : [ryan[1)] , taufiq[2)]]@informatika.lipi.go.id

Abstrak

Digital signature merupakan salah satu bagian dari metode kriptografi yang menggunakan fungsi hash untuk menandai sebuah data. Signature ini berguna untuk mengetahui bahwa data yang dikirim benar-benar berasal dari pengirim dan tidak mengalami perubahan. Sistem pencarian publikasi ilmiah berbasis semantik adalah suatu mesin pencari publikasi ilmiah yang menggunakan metadata berupa entry bibtex sebagai kriteria pencarian. Pada sistem tersebut, digital signature digunakan untuk identifikasi sumber data dan mencegah duplikasi data. Data bibtex, yang diperoleh dari penyedia layanan data publik seperti CiteSeer, Google Scholar, dll serta dari data milik sendiri, akan dikonversi ke bentuk metadata (dalam hal ini RDF) dan juga diproses menggunakan fungsi hashing dan enkripsi untuk membuat digital signature. Hasil pemrosesan ini kemudian disatukan ke dalam bentuk data RDF yang nantinya akan digunakan di repository metadata.

Kata Kunci : digital signature, kriptografi, bibtex, metadata.

PENDAHULUAN

Pada pengiriman data dan penyimpanan data, diperlukan suatu sistem keamanan untuk menjaga integritas dan keutuhan data. *Digital signature* merupakan salah satu bagian dari implementasi kriptografi dalam bidang keamanan untuk mengatasi problem tersebut. *Digital signature* merupakan sebuah fungsi satu arah, sehingga hasil dari fungsi ini tidak bisa dikembalikan ke bentuk semula. Hal ini berbeda dengan enkripsi yang berupa fungsi dua arah sehingga bisa dikembalikan ke bentuk semula. Proses ini, secara default, menggunakan fungsi hash dan enkripsi asimetrik untuk menghasilkan suatu kode dengan panjang tertentu (biasanya sekitar 512 bit).

Mesin pencari semantik adalah perkembangan dari mesin pencari biasa dimana hasil yang ditampilkan adalah dokumen yang sesuai dengan kriteria pencarian yang diinginkan. Dengan kata lain, suatu mesin pencari semantik adalah suatu mesin pencari yang mengambil "makna" kata sebagai salah satu faktor dalam algoritma pencariannya atau menawarkan suatu pilihan tentang makna kata atau frase kepada pengguna. [7]

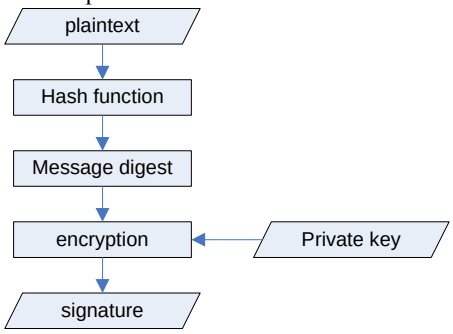
Dalam suatu sistem pencarian berbasis semantik dibutuhkan adanya suatu repositori metadata yang cukup handal sehingga proses pencarian dapat berjalan lebih cepat dan akurat. Repositori metadata sendiri adalah suatu database data tentang data (metadata) yang menyediakan sarana yang konsisten dan dapat diandalkan dalam mengakses data yang mana repositori tersebut dapat berupa lokasi fisik ataupun virtual dimana metadata diambil dari sumber lain.

Pada makalah ini dipaparkan penggunaan *digital signature* pada metadata untuk pencarian publikasi ilmiah berbasis semantik. Dalam sistem ini, *digital signature* digunakan untuk menvalidasi data yang diperoleh dari sumber data dan mencegah terjadinya duplikasi metadata yang masuk ke repositori. Repositori ini menggunakan ontologi sebagai skema semantik sehingga memungkinkan melakukan penalaran secara otomatis terhadap data dan repositori semantik bekerja dengan model data yang fleksibel dan umum sehingga memungkinkan untuk menafsirkan dan mengadopsi skema metadata atau ontologi yang terus berubah.[4]

DIGITAL SIGNATURE

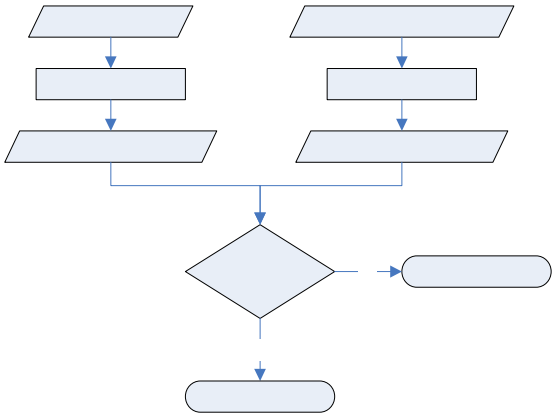
Digital signature, sebagaimana telah diuraikan sebelumnya, merupakan sebuah contoh dari metoda otentikasi untuk membuktikan kebenaran sebuah identitas. Sebuah data yang ditambahkan digital signature akan segera ketahuan apabila data tersebut mengalami perubahan/diubah. *Digital signature* diproses dengan cara meng-enkripsi sebuah message

digest menggunakan kriptografi asimetrik. Cara ini dapat membantu verifikasi identitas pengirim data. Pengirim memproses *message digest* dari data yang akan dia persiapkan untuk dikirim dan menandainya dengan enkripsi dari *message digest* ini. Hasil dari enkripsi *message digest* inilah yang disebut dengan *signature*[1][3]. Ilustrasi proses penandaan data ini diperlihatkan pada Gambar 1.



Gambar 1. Proses pembuatan *signature*.

Pengirim kemudian mengirimkan baik data asli maupun *digital signature* kepada penerima. Penerima memverifikasi *signature* dengan cara yang pertama men-dekripsi *signature* dengan kunci publik yang dimiliki pengirim. Hasil dari proses dekripsi ini berupa *message digest*. Di sisi lain, penerima juga melakukan fungsi hash terhadap data yang diterima. Kemudian, penerima membandingkan kedua *message digest* tersebut. Jika keduanya sama, penerima dapat memverifikasi bahwa data adalah sungguh-sungguh berasal dari pengirim dan tidak mengalami perubahan selama proses pengiriman.[3]. Proses ini ditunjukkan oleh diagram alir pada gambar 2.



Gambar 2. Proses verifikasi

Secara keseluruhan, proses pembuatan *digital signature* dikelompokkan menjadi 3 bagian utama [2], yaitu:

1. Data terenkripsi;
2. *Message Authentikasi Code*(MAC);
3. Fungsi Hash.

WEB SEMANTIK

Metadata

Metadata adalah data tentang data dimana ia dapat menangkap makna dari data tersebut. Saat ini, format isi Web lebih ditujukan pada pengguna (manusia) tetapi tidak bagi mesin. Manusia bisa mengenali makna yang tergantung pada isi Web tersebut tetapi bagi mesin, hal itu sangat menyulitkan. Pada Web Semantik dibutuhkan pengenalan atas makna (semantik) dari suatu isi sehingga mesin dapat memproses dengan lebih mudah [4]

Ontologi

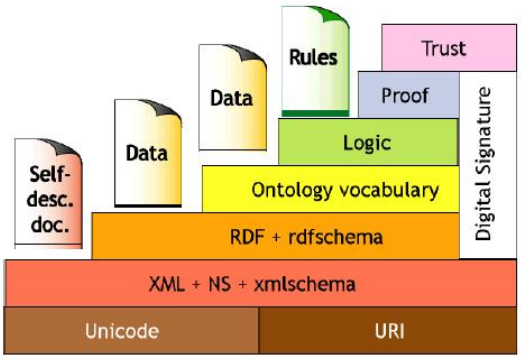
Ontologi pada awalnya merupakan cabang metafisika yang berkaitan dengan identifikasi secara umum jenis-jenis sesuatu yang secara nyata ada dan bagaimana mendeskripsikannya. Pada perkembangannya, istilah tersebut menjadi salah satu istilah yang digunakan dalam ilmu komputer dan diberikan arti teknis yang agak berbeda dengan makna asalnya. Devi dkk [5] mengumpulkan berbagai macam definisi tentang ontologi. Salah satunya yang dikemukakan oleh Gruber yang kemudian diperbaiki oleh Studer: ”*Ontologi adalah suatu spesifikasi yang spesifik dan formal dari suatu konseptualisasi*.”
Secara umum, ontologi mendeskripsikan secara formal domain pembicaraan. Biasanya suatu ontologi terdiri dari daftar istilah yang terbatas dan hubungan (*relationship)* yang ada antar istilah tersebut. Istilah-istilah itu menunjukkan konsep-konsep (kelas obyek) penting yang ada dalam domain tersebut. Sebagai contoh, dalam domain publikasi ilmiah, istilah penulis, penerbit, judul merupakan beberapa konsep penting dalam domain itu. Sedangkan hubungan biasanya meliputi hirarki antar kelas. Sebagai contoh, setiap artikel dalam prosiding adalah bagian dari prosiding tersebut. Selain hubungan hirarki, ontologi juga bisa memuat informasi seperti:

- properti
- pembatasan nilai
- pernyataan disjoin
- spesifikasi hubungan logik antar obyek

Dalam konteks Web, ontologi menyediakan suatu pemahaman bersama atas suatu domain dimana hal ini sangat dibutuhkan untuk mengatasi masalah perbedaan terminologi. Baik ada satu istilah dengan makna yang beragam ataupun ada banyak istilah dengan makna yang sama. Sehingga diperlukan suatu pemetaan terminologi tertentu ke ontologi bersama atau dengan mendefinisikan pemetaan langsung antar ontologi. Selain itu, ontologi juga berguna meningkatkan akurasi pencarian isi Web dimana mesin pencari dapat mencari ke halaman dengan konteks yang dimaksud alih-alih mengumpulkan semua halaman dimana kata kunci yang ambigu muncul.

Berdasarkan skema desain Web Semantik yang dikemukakan oleh Berners-Lee (Gambar 3), bahasa ontologi yang digunakan untuk Web saat ini:

- XML (*Extensible Markup Language*)
  Menyediakan sintak untuk luaran dokumen terstruktur, tetapi belum dipaksakan untuk menggunakan batasan semantik pada makna dokumen.
- XML Schema
  Bahasa untuk pembatasan struktur dari dokumen XML.
- RDF (*Resource Description Framework*)
  Model data untuk obyek dan hubungan antar obyek, menyediakan semantik yang sederhana untuk model data tersebut, dan dapat disajikan dalam sintak XML.
- RDF Schema
  Adalah kosakata untuk menjelaskan properti dan kelas dari sumber RDF, dengan semantik untuk hirarki penyamarataan dari properti dan kelas.
- OWL (*Ontology Web Language*)
  Perluasan kosakata untuk menjelaskan properti dan kelas, seperti hubungan antar kelas, kardinalitas, persamaan (*equality*) dan lain-lain.



Gambar 3. Layer Web Semantik [6]

Logika

Logika memberikan kita bahasa formal untuk mengekspresikan pengetahuan serta semantik formal yang dapat dipahami dimana makna kalimat dapat didefinisikan tanpa perlu mengoperasikan pengetahuan. Selain itu, penalar otomatis dapat menarik kesimpulan dari pengetahuan yang diberikan, sehingga membuat pengetahuan implisit menjadi eksplisit. Kelebihan logika adalah ia dapat memberikan penjelasan atas kesimpulan yang diambil dengan melacak balik rangkaian penarikan kesimpulan.

Agen

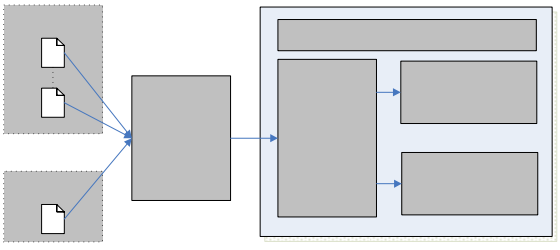
Agen merupakan software yang bekerja untuk kepentingan user secara otonom dan proaktif. Agen Web Semantik menggunakan teknologi yang sudah disebutkan sebelumnya seperti metadata yang digunakan untuk mengidentifikasi dan mengekstrak informasi dari Web, ontologi yang digunakan untuk membantu pencarian, menginterpretasikan informasi yang diperoleh dan berkomunikasi dengan agen lain. Demikian juga logika yang digunakan untuk memproses informasi dan menarik kesimpulan.

DESAIN SISTEM

Bibtex merupakan alat untuk memformat daftar referensi (bibliografi) dimana kita bisa memisahkan informasi bibliografi dari tampilan informasi tersebut dalam file yang terpisah. Suatu basis data Bibtex disimpan dalam format file berekstensi .bib yang merupakan file teks biasa sehingga mudah untuk dilihat dan diedit. Struktur yang dipakai cukup sederhana. Berikut contoh suatu isian Bibtex:

```
inproceedings{ petersen92dynamic,
  author = "Paul M. Petersen and David
A. Padua",
  title    =    "Dynamic    Dependence
Analysis: {A} Novel Method for Data
Dependence Evaluation",
  booktitle   =   "1992   Workshop   on
Languages and Compilers for Parallel
Computing",
  number = "757",
  publisher    =    "Berlin:    Springer
Verlag",
  address = "New Haven, Conn.",
  pages = "64--81",
  year = "1992",
  url="citeseer.ist.psu.edu/5218.html"
}
```

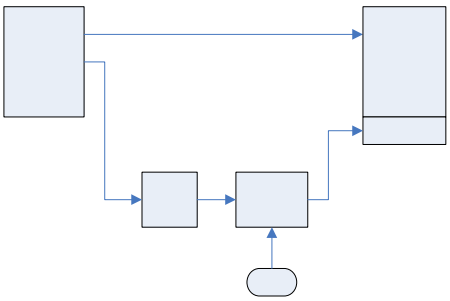
Bibtex tersebut akan memudahkan informasi tentang publikasi ilmiah dijadikan metadata dalam web semantik dengan cara mengkonversi struktur data bibliografi dalam Bibtex ke bentuk metadata yang digunakan dalam web semantik, dalam hal ini ke bentuk RDF. Metadata inilah yang selanjutnya akan diproses untuk menghasilkan *digital signature*. Pada gambar 4 ditunjukkan arsitektur sistem yang berkaitan dengan pemrosesan metadata.



Gambar 4. Arsitektur sistem untuk pemrosesan metadata

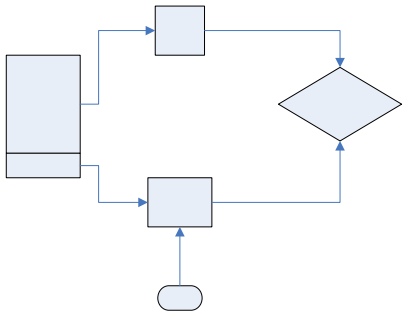
Data Bibtex diperoleh dari penyedia layanan data publik seperti CiteSeer[1], ACM[2], BibSonomy[3], Google Scholar[4] dll serta dari data milik sendiri. Data-data tersebut tersedia dalam bentuk file teks dalam format *.bib*. Modul pemrosesan data akan mengubah data tersebut menjadi format RDF. Hasil pemrosesan kemudian disatukan di modul integrasi data dan data inilah yang nantinya digunakan di repositori metadata. Ada 2 tool yang disediakan yaitu tool untuk memantau kapasitas penyimpanan dan tool untuk analisis data lampau dan pelaporan.

Pada modul pemrosesan data ini, proses pembuatan digital signature juga dilakukan. Metadata yang telah didapat kemudian diproses menggunakan fungsi hash dan menghasilkan *message digest*. Selanjutnya message digest ini di-enkripsi dengan menggunakan kunci *private KRa* untuk membentuk *signature*. *Signature* ini yang kemudian disertakan kembali bersama metadata aslinya dan dikonversi ke dalam format RDF. Gambar 5 menunjukkan proses pembangkitan *digital signature* ini.



Gambar 5. Pembuatan digital signature pada metadata

Tujuan dari dipasangkannya *digital signature* ini pada metadata adalah untuk identifikasi sumber data dan mencegah terjadinya duplikasi. Identifikasi sumber data dilakukan dengan mem-verifikasi RDF yang tersimpan dalam repositori. Proses verifikasi ini ditunjukkan oleh gambar 6.



Gambar 6. Proses verifikasi metadata

Metadata

[1] http://citeseer.ist.psu.edu/
[2] http://portal.acm.org/
[3] http://www.bibsonomy.org/
[4] http://scholar.google.com/

Proses verifikasi dilakukan dengan cara men-dekripsi signature menggunakan kunci publik *KUa* dan melakukan fungsi hash terhadap metadata. Kemudian, hasil dari kedua langkah tersebut dibandingkan. Jika hasilnya sama, maka metadata tersebut sesuai dengan aslinya tanpa mengalami perubahan. Jika tidak sama maka metadata tersebut telah diubah.

Sedangkan untuk mencegah duplikasi data, apabila terdapat 2 atau lebih isian bibtex dengan *signature* yang sama, maka sistem hanya akan mengambil salah satu saja dan menghapus isian bibtex yang lain.

IMPLEMENTASI

Rancangan repositori metadata menggunakan digital signature di atas diimplementasikan dengan Java Web Service (Java 2 Standard Edition 1.5.0_07 yang berjalan di atas server web Apache Tomcat 5.5.20. Fungsi hash yang digunakan adalah MD5 dengan enkripsi RSA. Contoh metadata yang digunakan yang merupakan hasil konversi dari file Bibtex sebagai berikut:

```
<?xml version='1.0' encoding='ISO-
8859-1'?>
<rdf:RDF

xmlns:bibtex="http://www.informatika.l
ipi.go.id/bibtex#"

xmlns:rdf="http://www.w3.org/1999/02/2
2-rdf-syntax-ns#"

xmlns:vcard="http://www.w3.org/2001/vc
ard-rdf/3.0#"

xmlns:dc="http://purl.org/dc/elements/
1.1/"

xmlns:dct="http://purl.org/dc/terms/"

xmlns:rdfs="http://www.w3.org/2000/01/
rdf-schema#">

  <bibtex:InProceedings
rdf:about="petersen92dynamic">
    <dct:isPartOf>
      <bibtex:Proceedings
rdf:about="petersen92dynamic_1992_Work
shop_on_Languages_and_Compilers_for_Pa
rallel_Computing"/>
    </dct:isPartOf>
    <bibtex:pages>64-81</bibtex:pages>
    <dc:title>Dynamic Dependence
Analysis: A Novel Method for Data
Dependence Evaluation</dc:title>

<dc:identifier>citeseer.ist.psu.edu/52
18.html</dc:identifier>
    <dc:creator>
      <rdf:Seq>
        <rdf:li>
```

Metadata

Sig

Hash Enkripsi

```
          <bibtex:Person
rdf:about="petersen92dynamic_Petersen_
Paul_M_"/>
        </rdf:li>
        <rdf:li>
          <bibtex:Person
rdf:about="petersen92dynamic_Padua_Dav
id_A_"/>
        </rdf:li>
      </rdf:Seq>
    </dc:creator>
  </bibtex:InProceedings>
  <bibtex:Proceedings
rdf:about="petersen92dynamic_1992_Work
shop_on_Languages_and_Compilers_for_Pa
rallel_Computing">
    <dc:date>1992</dc:date>
    <dc:publisher>
      <bibtex:Organization
rdf:about="petersen92dynamic_Berlin__S
pringer_Verlag"/>
    </dc:publisher>
    <dc:title>1992     Workshop     on
Languages and Compilers for Parallel
Computing</dc:title>
  </bibtex:Proceedings>
  <bibtex:Person
rdf:about="petersen92dynamic_Padua_Dav
id_A_">
    <vcard:N rdf:parseType="Resource">
      <vcard:Given>David</vcard:Given>

<vcard:Family>Padua</vcard:Family>
      <vcard:Other>A.</vcard:Other>
    </vcard:N>
    <vcard:FN>David                  A.
Padua</vcard:FN>
  </bibtex:Person>
  <bibtex:Person
rdf:about="petersen92dynamic_Petersen_
Paul_M_">
    <vcard:FN>Paul                   M.
Petersen</vcard:FN>
    <vcard:N rdf:parseType="Resource">
      <vcard:Given>Paul</vcard:Given>
      <vcard:Other>M.</vcard:Other>

<vcard:Family>Petersen</vcard:Family>
    </vcard:N>
  </bibtex:Person>
  <bibtex:Organization
rdf:about="petersen92dynamic_Berlin__S
pringer_Verlag">
    <vcard:FN>Berlin:          Springer
Verlag</vcard:FN>
  </bibtex:Organization>
  <bibtex:SourceFile
rdf:about="bibfile">

<bibtex:absolutePath>file:/D:/MyDocs/m
yResearch/2008/Semantic.Web/InWork/RDF
%20Converter/bibtex2rdf/../../Data/bib
/sriti.bib</bibtex:absolutePath>
  </bibtex:SourceFile>
  <bibtex:Signature>
    8ó¾ÃQuÞ118õ111ÚQaa!!uóaóQõó#1aõµ
  </bibtex:Signature>
    <rdf:Seq rdf:about="referenceList">
      <rdf:li
  rdf:resource="petersen92dynamic"/>
    </rdf:Seq>
  </rdf:RDF>
```

Metadata hasil konversi tersebut di-*load* ke repositori dan siap digunakan melalui tool manajemen. Sampai saat ini implementasi baru sampai tool manajemen sedangkan layer penyimpanan dan inferensi belum dilakukan.

KESIMPULAN

Telah dipaparkan rancangan penggunaan digital signature pada metadata yang akan digunakan pada sistem pencarian publikasi ilmiah berbasis semantik. Publikasi ilmiah yang digunakan disusun menggunakan format BibTex yang dikonversi ke format RDF yang kemudian diproses menggunakan fungsi hash dan enkripsi untuk menghasilkan digital signature dan disimpan di repositori. Digital signature ini digunakan untuk identifikasi sumber data dan mencegah duplikasi. Pengembangan lebih lanjut masih diperlukan untuk mendapatkan sistem yang efisien, cepat, akurat, dan aman.

DAFTAR PUSTAKA

[1] Zhiqun Chen (2000). Java Card Technology for Smart Cards: Architecture and Programmer's Guide. Sun Microsystem, Inc.

[2] William Stallings (1999). Cryptography and Network Security: Principles and Practice, Prentice Hall, Inc

[3] Nova Hadi Lestriandoko dan Rifki Sadikin (2008), Design of Smartcard's Mutual Authentication Using Zero Knowledge Protocol, Proceedings of 4th International Conference on Information Communication Technology and Systems, Volume 1, Number 1.

[4] Taufiq Wirahman dan Devi Munandar (2008), Repository Metadata dan Ontologi pada Pencarian Publikasi Ilmiah Berbasis Semantik, Proceeding Seminar Nasional Riset Teknologi Informasi 2008, Volume 3, 2008.

[5] Devi Munandar, Nurhayati Masthurah, Taufiq Wirahman (2007), Pemetaan Ontologi Publikasi Ilmiah Dalam Mendukung Web Semantik, Prosiding Seminar Nasional Teknologi Industri 2007.

[6] Grigoris Antoniou, Frank van Harmelen (2004), A Semantic Web Primer, MIT Press.

[7] Phill Midwinter, 2007, Is Google A Semantic Search Engine, http://www.grantmidwinter.com