PENERAPAN *SEMANTIC SIMILARITY* PADA KRIPTOGRAFI SUATU DOKUMEN TEKS DALAM BAHASA INDONESIA

***Detty Purnamasari*[1]**
***I Wayan Simri Wicaksana*[2]**
***Anissa Lintang R.*[3]**
***Anneke Annassia P.S.*[4]**
***Hendry Gustin*[5]**

[1]*Jurusan Sistem Informasi, FIKTI, Universitas Gunadarma*
[2] *Program Doktor Teknologi Informasi, Universitas Gunadarma*
[3,4,5]*Jurusan Teknik Informatika, FTI, Universitas Gunadarma*
[1]*detty@staff.gunadarma.ac.id,* [2]*iwayan@staff.gunadarma.ac.id,*
[3]*thebulletstring@gmail.com,* [4]*nekkeps@gmail.com,* [5]*henzstyle12@gmail.com*

ABSTRAK

Penerapan kriptografi digunakan untuk memberikan keamanan pada dokumen penting. Kriptografi terdiri dari proses ekripsi dan deskripsi. Enkripsi suatu dokumen teks telah banyak diterapkan dengan berbagai macam metode untuk mengubah dokumen asli menjadi dokumen yang terkunci, setiap metode/algoritma memiliki tingkat kerumitan yang berbeda. Namun, penerapan enkripsi yang terlalu mencolok dapat mengundang hacker untuk berusaha memecahkan enkripsi tersebut. Artikel ini mengembangkan pendekatan untuk penerapan enkripsi pada kriptografi suatu dokumen teks dalam bahasa Indonesia dengan mengenkripsi dokumen teks asli ke bentuk dokumen teks yang mengandung kata/kalimat lain dengan tingkat similarity tertentu. Penerapan enkripsi ini, diharapkan menjadi lebih efektif dan tidak menarik perhatian hacker, karena pada penelitian ini enkripsi yang dihasilkan masih berbentuk teks dokumen utuh, tetapi memiliki makna yang berbeda dengan dokumen aslinya. Pada artikel ini dibuat ilustrasi untuk pendekatan yang dikembangkan untuk proses enkripsi.

***Kata Kunci**: dokumen, enkripsi, kriptografi, semantic similarity, teks*

PENDAHULUAN

Saat ini, tindakan pencurian atau penyelundupan dalam dunia Internet sudah tidak asing lagi.

Mulai dari hal yang sederhana seperti membajak sosial media, sampai penyadapan suatu data pada dunia bisnis bahkan pemerintahan. Metode dan sistem keamanan dibutuhkan untuk menjaga data yang dimiliki oleh masing-masing orang. Pada artikel ini fokus membicarakan tentang pengamanan dari pencurian dan penyadapan data/dokumen teks. Salah satu cara pengamanannya yaitu dengan menggunakan Kriptografi. Kriptografi adalah ilmu dan seni untuk menjaga kerahasiaan berita [Scheier, 1996]. Definisi lainnya yaitu; ilmu yang

mempelajari teknik-teknik matematika yang berhubungan dengan aspek keamanan informasi seperti kerahasiaan data, keabsahan data, integritas data, serta autentikasi data [Menezes, 1996].

Pada kriptografi, terdapat dua istilah utama yaitu Enkripsi dan Dekripsi. Enkripsi adalah proses mengamankan suatu informasi dengan membuat informasi tersebut tidak dapat dibaca tanpa bantuan pengetahuan khusus, sedangkan Dekripsi adalah kebalikan dari enkripsi yaitu proses mengkonversi data yang sudah di enkripsi kembali menjadi data aslinya, sehingga dapat dibaca atau dimengerti kembali. Dua kunci yang digunakan pada kriptografi, yaitu: kunci publik yang digunakan untuk enkripsi dan kunci privat yang digunakan untuk dekripsi [Goyal, 2012].

Penelitian yang pernah dilakukan oleh Jamgekar et.al. (2013) untuk keamanan adalah dalam hal keamanan transmisi file dengan menggunakan algoritma RSA yang telah di modifikasi yang merupakan kunci kriptografi asymmetric yang disebut dengan kunci public [Jamgekar, 2013]. Keamanan dokumen teks juga dapat dilakukan dengan menggunakan text watermarking yang mengkombinasikan gambar serta teks untuk mengenkripsi dokumen, dan cara ini pernah dilakukan oleh Jaseena et.al. [Jaseena, 2011].

Pada pengembangannya, sudah banyak metode dan algoritma dari kriptografi. Selain algoritma RSA yang dapat digunakan untuk enkripsi, pendekatan yang dapat dikombinasikan ke dalam metode enkripsi adalah pendekatan *semantic similarity*. Semantik *Similarity* adalah metode untuk melakukan pencarian kesepadanan makna dari konsep/kata. Semantik menyediakan aturan untuk menafsirkan sintaks yang tidak memberikan makna secara langsung tetapi membatasi kemungkinan penafsiran dari apa yang dinyatakan [Euzenat, 2001].

Artikel ini mengembangkan pendekatan untuk penerapan semantik *similarity* pada enkripsi yang merupakan bagian dari kriptografi pada suatu dokumen teks dengan Bahasa Indonesia. Pendekatan yang dikembangkan menghasilkan dokumen teks baru yang dihasilkan dan memiliki arti yang berbanding terbalik dari dokumen aslinya, sehingga metode ini dapat membantu pengamanan dari suatu dokumen teks.

PENDEKATAN ENKRIPSI DENGAN *SEMANTIC SIMILARITY*

Pendekatan yang dikem-bangkan pada penelitian ini adalah melakukan enkripsi pada dokumen teks dalam Bahasa Indonesia, sehingga struktur kalimat yang diperhatikan pada pendekatan ini. Dokumen teks merupakan rangkaian paragraf yang terdiri dari banyak kalimat. Kalimat itu sendiri adalah rangkaian kata yang dapat mengungkapkan gagasan, pikiran, atau perasaan. Kalimat merupakan satuan bahasa terkecil yang mengungkapkan pikiran yang utuh, baik dengan cara lisan maupun tulisan. Pada kalimat sekurang-kurangnya ha-rus memiliki subjek (S) dan predikat (P). Bila tidak memiliki subjek dan predikat, maka bukan disebut kalimat tetapi disebut frasa. Unsur kalimat dalam Bahasa Indonesia adalah Subjek (S), Predikat (P), Objek (O), dan Keterangan (K).

Unsur kalimat telah diketa-hui, sehingga dapat dilakukan *preprocessing*, yaitu dengan membuat database yang berisi i). preposisi, ii). subjek, iii).objek, dan iv).keterangan, serta v). predikat. Gambar 1 adalah proses enkripsi yang dilakukan pada pendekatan di artikel ini. Tahapan enkripsi adalah pem-bacaan dokumen, pemisahan kalimat,

pemisahan struktur kalimat, operasi *semantic*, dan *rewriting*.

i. Pembacaan Dokumen

Dokumen yang digunakan pada pendekatan enkripsi ini
adalah dokumen yang berupa teks seluruhnya tanpa memiliki gambar, grafik, ataupun tabel didalamnya. Selain itu, dokumen menggunakan Bahasa Indonesia, serta penulisannya sesuai dengan EYD (Ejaan Yang Disempurnakan).

ii. Pemisahan Kalimat

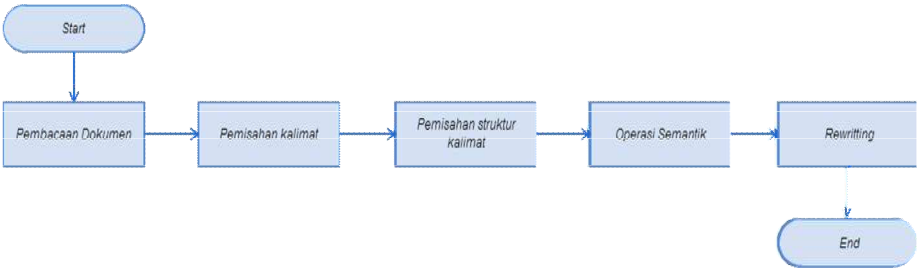
Pemisahan kalimat dapat dilakukan dengan metode berikut:

a. Posisi pembacaan dokumen dimulai dari karakter awal dokumen.
b. Jika pembacaan dokumen menemukan karakter tanda baca selain titik (“.”) maka karakter tersebut akan dihapus, dan jika pembacaan dokumen menemukan karakter titik (“.”) atau tanda tanya (?) atau tanda seru (!) , maka pisahkan kalimat dari posisi awal pembacaan dokumen hingga karakter titik/tanda Tanya/tanda seru yang ditemukan dan simpan ke database sementara.
c. Setelah memisahkan kalimat, maka pindahkan posisi awal pembacaan dokumen pada karakter awal dari kalimat selanjutnya.
d. Lakukan kembali langkah b)dan langkah c). hingga akhir dokumen atau hingga tidak ditemukan lagi karakter di dalam dokumen.

iii. Pemisahan Struktur Kalimat

Sebelum melanjutkan lang-kah selanjutnya terlebih dahulu dilakukan penghilangan tanda baca pada setiap kalimat. Setiap kalimat yang telah disimpan ke dalam database kemudian dila-kukan pemisahan struktur kalimat, sehingga didapatkan subjek, predikat, objek, keterangan, dan preposisi yang dimasukkan kembali ke dalam database yang berbeda. Pemisahan struktur kalimat dilakukan menggunakan langkah-langkah di bawah ini:

a. ***Subjek***, Pembacaan karakter selalu dimulai di awal karakter, sehingga setiap karakter dari awal hingga bertemu karakter spasi dan karakter bukanlah huruf ka-pital akan dimasukkan ke dalam struktur subjek. (kata pertama)
b. ***Predikat***, Karakter selanjutnya setelah mene-mukan subjek hingga bertemu karakter spasi akan dikate-gorikan sebagai struktur predikat. (kata kedua)
c. Karakter selanjutnya setelah ditemukannya subjek dan predikat (karakter sebagai kata ketiga) hingga menemukan karakter spasi akan dicari keberadaannya di dalam ‘database preposisi’. Apabila ditemukan, maka kata tersebut akan dikate-gorikan ke dalam kategori ***preposisi*** dan kata setelah preposisi adalah ***keterangan***. Tetapi, apabila kata tersebut tidak ditemukan di dalam ‘database preposisi’, maka kata tersebut merupakan ***objek*** dari kalimat, dan kembali melakukan pembaca-an karakter.
d. Karakter selanjutnya (sebagai kata keempat) hingga tidak ditemukan lagi karakter akan dimasukkan ke dalam kate-gori keterangan.



Gambar 1. Tahapan Proses Enkripsi Dokumen Teks

iv. Operasi *semantic*

Struktur kalimat telah dipisahkan, langkah selanjutnya adalah menentukan kata ganti untuk setiap struktur kalimat yang telah ditemukan pada langkah pemisahan struktur kalimat, sesuai dengan jenis struktur kalimatnya. Berikut ini adalah langkah yang dilakukan untuk mengubah dokumen asli menjadi dokumen yang sudah terenkripsi:

a. Ketika kata subjek, objek, dan keterangan memiliki semantik *similarity* dengan kata lain (bukan nama), maka kata tersebut kemudian dilakukan operasi semantik. Namun, apabila subjek, objek, ataupun keterangan tidak memiliki *similarity* apapun dengan kata lain (merupakan nama), maka kata tersebut akan dilakukan penggantian kata dari daftar kata yang terdapat di dalam 'database nama' menggunakan rumus:
   X = n + 2…….(1)
   Dimana n adalah kata awal pada kalimat yang akan di enkripsi.Setelah kata ganti sudah ditentukan, maka seluruh kata tersebut yang berada di dalam dokumen diganti dengan kata ganti tersebut.
b. Setiap predikat pasti memiliki *similarity* dengan kata kerja lainnya. Predikat kemudian dilakukan operasi semantik.
c. Preposisi yang terdapat pada setiap kalimat diubah menjadi preposisi antonimnya.

Berikut langkah-langkah operasi semantik predikat:

1) Hal pertama yang dilakukan adalah mencari tingkat sinonim dari kata kerja tersebut, tanpa menghilangkan imbuhan. Pencarian sinonim kata menggunakan:
   http://www.artikata.com
2) Mendaftarkan seluruh sinonim dari predikat tersebut yang memiliki tingkatan level pertama ke dalam database predikat.
3) Menerjemahkan predikat dan setiap sinonimnya ke dalam Bahasa Inggris dan memasukkannya ke dalam database.
4) Mencari bobot *similarity* antara predikat dan setiap sinonimnya, pada wordnet *similarity*: http://www.marimba.d.umn.edu/cgi-bin/*similarity*/*similarity*.cgi, dan menggunakan metode Jiang & Conrath [Budanitsky, 2006], karena perhitungan *similarity* dengan *Jiang & Conrath (JNC)* menunjukkan hasil yang paling baik [Sridhara, 2008] untuk mencari bobot *similarity* sebagai nilai untuk mengukur kesepadanan kata.
5) Menentukan kata ganti untuk predikat berdasarkan sinonim yang memiliki angka *similarity* terbesar.

6) Menerjemahkan sinonim yang telah terpilih ke dalam Bahasa Indonesia.
7) Mengganti predikat dengan sinonim yang terpilih.

v. Rewriting

Setelah operasi semantik dan penggantian kata telah dilakukan, maka langkah selanjutnya adalah menggabung-kan kembali setiap kata sesuai struktur sebelumnya kemudian menggabungkan setiap kalimat menjadi paragraf seperti urutan paragraf sebelumnya, sehingga diperoleh hasil enkripsi dari dokumen tersebut.

vi. Hasil

Hasil enkripsi berupa dokumen teks yang baru dengan struktur kalimat, urutan paragraf, serta jumlah paragraf yang sama seperti dokumen yang sebenarnya.

vii. Proses Deskripsi

Kriptografi tidak akan memiliki nilai jika dokumen yang sudah dienkripsi tidak bisa dikembalikan menjadi dokumen yang sebenarnya. *Decryption Processing* atau proses dekripsi dilakukan menggunakan tahapan yang sama seperti proses enkripsi dengan membedakan rumus yang digunakan(misalnya dengan rumus X**,** maka rumus ini yang digunakan untuk deskripsi). Hasil dari proses dekripsi ini menghasilkan dokumen baru yang berisi data seperti yang terdapat di dalam dokumen sebenarnya.

ILUSTRASI

Berikut ini merupakan ilustrasi dari penerapan *semantic similarity* pada proses enkripsi untuk dokumen teks berbahasa Indonesia.

1. Pembacaan Dokumen

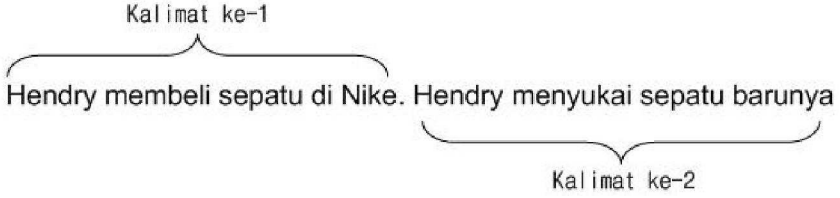
Pada contoh dokumen teks (Gambar 2) adalah dokumen yang terdiri dari 2 (dua) kalimat.

2. Pemisahan Kalimat

Awal kalimat dalam bahasa Indonesia diawali dengan menggunakan huruf capital, sehingga karakter ini merupakan karakter awal, sampai ditemukannya karakter titik (.)atau tanda tanya (?) atau tanda seru (!) yang merupakan tanda akhir kalimat. Tampak pada Gambar 3 adalah ilustrasi untuk pemisahan kalimat. Pada Gambar 3, pembacaan awal karakter adalah huruf capital (huruf “H”), maka ini adalah awal kalimat. Kemudian ditemukan pembacaan tanda titik (.), ini merupakan tanda dari akhir kalimat, sehingga didapatkan 2 kalimat, yaitu:

Kalimat ke-1: Hendry membeli sepatu di Nike

Kalimat ke-2: Hendry menyukai sepatu barunya.



Gambar 2. Contoh Dokumen Teks

Gambar 3. Ilustrasi Pemisahan Kalimat

3. Pemisahan Struktur Kalimat

Pada pemisahan struktur kalimat, masing-masing kalimat yang sudah ditemukan pada tahap 2 diatas akan dipecah per kata. Pada Gambar 4. dibuat ilustrasi untuk menentukan struktur kalimat ke-1 (Hendry membeli sepatu di nike). Pembacaan karakter dmulai dari karakter kesatu sampai ditemukannya spasi, maka kata "Hendry" adalah kata ke-1 sebagai 'Subjek'. Pembacaan karakter dilanjutkan kembali dan ditemukan kata ke-2 "membeli" sebagai 'Predikat', dan pada kata ke-3 "sepatu" dilakukan pencarian pada 'database preposisi', dan hasilnya tidak ditemukan, maka kata ke-3 adalah 'Objek'. Kata ke-4 'di' dicari pada 'database preposisi', dan hasilnya ditemukan, maka kata ke-4 adalah preposisi, dan kata ke-5 "Nike" merupakan 'Keterangan' karena berada setelah preposisi. Berikut rincian struktur kalimat untuk kalimat ke-1: Hendry = Subjek; membeli = Predikat; sepatu = Objek; di = Preposisi; Nike = Keterangan.

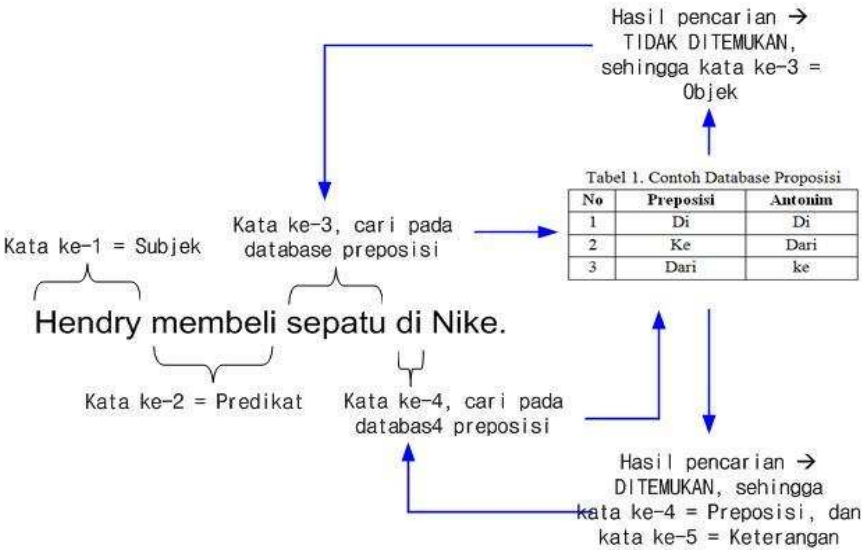
4. Operasi Semantik

Setiap struktur kalimat yang sudah ditemukan kemudian dilakukan operasi semantik sesuai dengan jenis struktur kalimatnya. Berikut ini ilustrasi langkah operasi *semantic* untuk contoh kalimat ke-1 (Hendry membeli sepatu di Nike):

Subjek

Subjek pada kalimat ke-1 adalah "Hendry", kemudian kata ini dicari pada 'database subjek' untuk menentukan kata "Hendry" ada pada posisi record ke berapa. Setelah ditemukan no record, kemudian dengan menggunakan rumus X, subjek 'Hendry' akan diganti dengan subjek yang berada pada nomer record berdasarkan rumus X tersebut

.

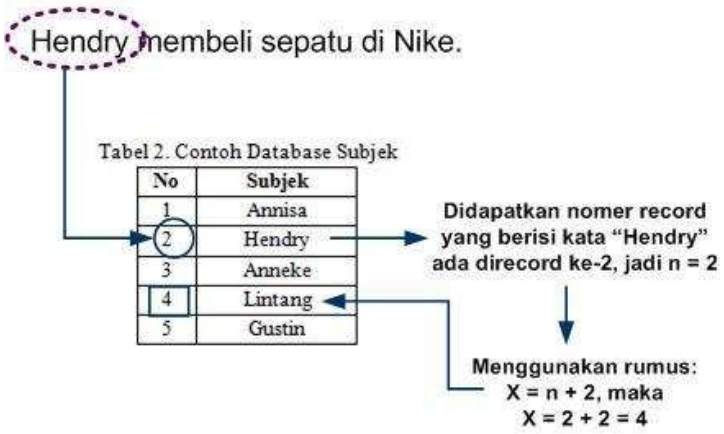


Gambar 4. Ilustrasi Pemisahan Struktur Kalimat

Misalkan, ditentukan **rumus x = n + 2**, dengan variable x adalah nomer record baru sebagai posisi record untuk subjek pada dokumen enkripsi, dan variable n adalah nomer record dari subjek dokumen asli yang ditemukan pada 'database subjek'. Pada Gambar 5, subjek "Hendry" dicari pada 'database subjek', dan ditemukan nomer record = n = 2. Berdasarkan pada rumus yang digunakan, pada contoh adalah x = n + 2, maka nomer record baru = x = 2 + 2 = 4, maka subjek pada dokumen yang terenkripsi menjadi "**Lintang**".

Predikat

Predikat yang ditemukan pada kalimat ke-1 (Hendry membeli sepatu di Nike) adalah kata "membeli".(Lihat gambar 6).



Gambar 5. Ilustrasi Operasi untuk Subjek

Tabel 5. Contoh Database Predikat

No.	Kata	Sinonim	Bobot Similarity
1	Membeli (Buy)	Mengambil (Take)	0.2323
		Menukar (Exchange)	0.107
		Berbelanja (Shop)	0.1331
		Memperoleh (Obtain)	0.2452

Gambar 6. Ilustrasi Operasi Semantik untuk Predikat

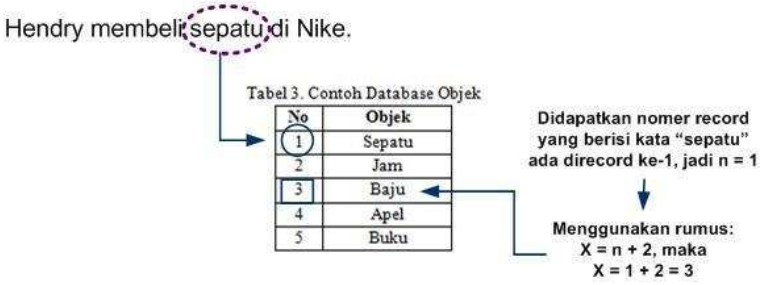
Kemudian dengan menentukan bobot *similarity* yang diinginkan, dicari *semantic* dari kata 'membeli'. Ilustrasi tampak pada Gambar 6. Misalkan pada ilustrasi tersebut, bobot *similarity* yang diinginkan pengguna adalah lebih besar atau sama dengan 0.2 (BS >= 0.2). Berdasarkan pada bobot *similarity* tersebut, ditemukan kata yang mempunyai keserupaan makna dengan 'membeli' pada bobot *semantic similarity* >= 0.2 adalah "mengambil", maka pada dokumen yang terenkripsi predikat akan berisi kata "**mengambil**".

Objek

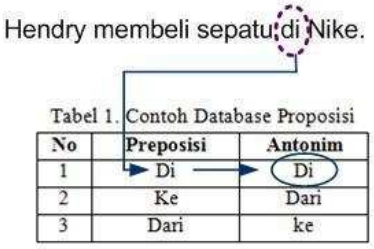
Kata ke-3 yang ditemukan pada kalimat ke-1 "sepatu", pada proses penentuan struktur kalimat dinyatakan sebagai objek, maka untuk mencari objek yang akan digunakan di dokumen yang sudah terenkripsi sama dengan tahap subjek. Ilustrasi tampak pada Gambar 7, dan dimisalkan rumus yang digunakan pada objek sama dengan rumus subjek: **x = n + 2**. Kata "sepatu" dicari pada 'database objek', dan ditemukan berada pada posisi record ke-1, sehingga dengan menggunakan rumus x = n + 2 = 1 + 2 = 3, maka posisi record ke-3 yaitu kata "**baju**" yang akan digunakan sebagai objek pada dokumen yang terenkripsi.

Preposisi

Kata ke-4 pada contoh kalimat ke-1 (Hendry membeli sepatu di Nike), yaitu kata 'di', kemudian dilakukan pencarian daftar antonimnya pada 'database preposisi', dan hasil yang didapatkan adalah kata yang sama yaitu "di". Ilustrasi pada Gambar 8.



Gambar 7. Ilustrasi Operasi untuk Objek



Tabel 1. Contoh Database Proposisi

No	Preposisi	Antonim
1	Di	Di
2	Ke	Dari
3	Dari	ke

Gambar 8. . Ilustrasi Operasi untuk Preposisi

Keterangan

Seperti halnya subjek dan objek, kata ke-5 pada contoh yaitu kata "Nike" yang merupakan struktur keterangan, kemudian dicari kata gantinya untuk dokumen yang terenkripsi pada "database keterangan", dan dengan menggunakan cara yang sama pada subjek dan objek, kata keterangan ini juga menggunakan suatu rumus, misalkan dengan rumus: x = n + 2. Pada Gambar 9, tampak ilustrasi untuk operasi enkripsi pada kata yang sebagai keterangan. Berdasarkan pada rumus yang sudah ditentukan oleh pengguna, maka kata "**Mall**" akan menggantikan "Nike" pada dokumen yang terenkripsi.

Hasil

Berdasarkan pada ilustrasi yang telah dicontohkan pada kalimat ke-1, maka hasil dokumen yang sudah terenkripsi adalah:

Dokumen Asli :
Hendry membeli sepatu di Nike.

Dokumen Terenkripsi :
Lintang mengambil baju di mall.

SIMPULAN DAN SARAN

Pendekatan yang dikem-bangkan pada artikel ini merupakan salah satu pendekatan yang dapat digunakan untuk melakukan enkripsi pada dokumen teks. Pendekatan ini merupakan salah satu cara yang dapat dipilih untuk melakukan keamanan pada dokumen teks dan memberikan cara yang tidak akan mengundang perhatian dari pihak yang tidak berwenang, karena masih dalam bentuk kalimat yang tetap memper-tahankan struktur kalimat. Pada penelitian selanjutnya, akan dikembangkan pendekatan enkripsi tetap dengan peman-faatan *semantic similarity*, tetapi akan dibahas lebih mendalam untuk stuktur kalimat, karena pada dokumen teks berbahasa Indonesia masih banyak peng-guna yang tidak menggu-nakan struktur kalimat SPOK. Selain itu, akan dikembangkan metode untuk memperkaya database yang digunakan pada pendekat-an enkripsi dengan *semantic similarity* ini.

Gambar 9. Ilustrasi Operasi untuk Keterangan

DAFTAR PUSTAKA

Budanitsky, A., Hirst, G. 2006 "Evaluating WordNet-based Measures of Lexical Semantic Relatedness" *Journal Computational Linguistics*, Vol. 32, page 13-47.

Euzenat, J. 2001 "Towards a principled approach to *semantic* interoperability" INRIA Rhône-Alpes, Montbonnot Saint-Martin (France)

Goyal, S. 2012 "A Survey on the Application of Cryptography" International Journal of Science and Technology, Vol. 1 No. 3.

Jamgekar, R.S., Joshi, G.S. 2013 "File Encryption and Decryption Using Secure RSA" *International Journal of Emerging Science and Engineering*, Vol. 1, Issue 4.

Jaseena K.U., John, A. 2011 "Text Watermarking using Combined Image and Text for Authentication and Protection" International Journal of Computer Applications, Vol. 20, No. 4

Menezes, A.J., Vanstone, S.A., Oorscot, P.C.V. 1996. *Handbook of Applied Cryptography* CRC Press, Massachusetts.

Scheier, B. 1996 *Applied Cryptography* John Wiley & Sons, Illinois.

Sridhara, G., Hill, E. Pollock, L., Vijay, S. K. 2008 "Identifying Word Relations in Software: A Comparative Study of Semantic similarity Tools" *Proceedings of the 2008 The 16th IEEE International Conference on Program Comprehension*, Page 123-132.