BAB II

LANDASAN TEORI

Untuk mendukung penelitian ini, maka perlu dikemukakan hal-hal atau teori-teori yang berkaitan dengan permasalahan dan ruang lingkup pembahasan sebagai landasan dalam pembuatan penelitian ini

2.1. Konsep Dasar Sistem

Berikut ini adalah beberapa pengertian dari sistem menurut para ahli yang dipaparkan berdasarkan sumbernya.

Sistem dapat didefinisikan dengan pendekatan prosedur dan pendekatan komponen, sistem dapat didefinisikan sebagai kumpulan dari prosedur-prosedur yang mempunyai tujuan tertentu.[2, p.34]

Sistem adalah setiap sesuatu terdiri dari obyek-obyek, atau unsur-unsur, atau komponen-komponen yang bertata kaitan dan bertata hubungan satu sama lain, sedemikian rupa sehingga unsur-unsur tersebut merupakan satu kesatuan pemrosesan atau pengolahan yang tertentu [5, p.152]

Berdasarkan dari pendapat dari berbagai ahli yang telah dipaparkan di atas dapat di tarik kesimpulan bahwa pengertian dari sistem adalah kumpulan dari komponen – komponen yang saling berhubungan dan saling bergantung satu dengan yang lain demi mencapai tujuan tertentu.

2.1.1. Elemen Sistem

Elemen Sistem adalah suatu sistem terdiri dari sejumlah elemen yang saling berinteraksi, yang artinya saling kerjasama membentuk satu kesatuan.[4. p.4]

mengenai elemen-elemen yang membentuk sebuah sistem :

1. Tujuan

Setiap sistem memiliki tujuan (Goal), entah hanya satu atau mungkin banyak. Tujuan inilah yang menjadi pemotivasi yang mengarahkan sistem.

2. Masukan

Masukan (input) sistem adalah segala sesuatu yang masuk ke dalam sistem dan selanjutnya menjadi bahan yang diproses.

3. Proses

Proses merupakan bagian yang melakukan perubahan atau transformasi dari masukan menjadi keluaran yang berguna dan lebih bernilai, misalnya berupa informasi dan produk, tetapi juga bisa berupa hal-hal yang tidak berguna, misalnya saja sisa pembuangan atau limbah.

4. Keluaran

Keluaran (output) merupakan hasil dari pemrosesan. Pada sistem informasi, keluaran bisa berupa suatu informasi, saran, cetakan laporan, dan sebagainya.

5. Batas

Yang disebut batas (boundary) sistem adalah pemisah antara sistem dan daerah di luar sistem (lingkungan). Batas sistem menentukan konfigurasi, ruang lingkup, atau kemampuan sistem.

1. Mekanisme Pengendalian dan Umpan Balik

Mekanisme pengendalian (control mechanism) diwujudkan dengan menggunakan umpan balik (feedback), yang mencuplik keluaran.

2. Lingkungan

Lingkungan adalah segala sesuatu yang berada di luar sistem. Lingkungan bisa berpengaruh terhadap operasi sistem dalam arti bisa merugikan atau menguntungkan sistem itu sendiri.

2.1.2. Karakteristik Sistem

Suatu sistem mempunyai karakteristik. Karakteristik sistem adalah sebagai berikut[3. p,54]:

1. Suatu sistem mempunyai komponen-komponen sistem (*components*) atau subsitem-subsistem. Komponen-komponen sistem tersebut dapat berupa suatu bentuk subsistem. Setiap subsistem memiliki sifat dari sistem yang menjalankan suatu fungsi tertentu dan mempengaruhi proses sistem secara keseluruhan.

2. Suatu sistem mempunyai batas sistem(*Boundary*)

   Ruang lingkup sistem merupakan daerah yang membatasi antara sistem dengan sistem yang lain atau sistem dengan lingkungan luarnya. Batasan sistem ini memungkinkan suatu sistem dipandang sebagai satu kesatuan yang tidak dapat dipisahkan.

3. Suatu sistem mempunyai lingkungan luar (*Environment*)

   Bentuk apapun yang ada diluar ruang lingkup atau batasan sistem yang mempengaruhi operasi sistem tersebut disebut lingkungan luar sistem. Lingkungan luar sistem ini dapat bersifat menguntungkan dan dapat juga bersifat merugikan sistem tersebut.

4. Suatu sistem mempunyai penghubung(*Interface*)

   Media yang menghubungkan sistem dengan subsistem yang lain disebut penghubung sistem atau interface. Penghubung ini memungkinkan sumber-sumber daya mengalir dari satu subsistem ke subsistem lain. Bentuk keluaran dari satu subsistem akan menjadi masukan untuk subsitem lain melalui penghubung tersebut. Dengan demikian, dapat terjadi suatu integrasi sistem yang membentuk satu kesatuan.penghubung,

5. Suatu sistem mempunyai tujuan (*goal*).

   Suatu sistem mempunyai tujuan dan sasaran yang pasti dan bersifat *deterministik*. Kalau suatu sistem tidak memiliki sasaran maka

operasi sistem tidak ada gunanya. Suatu sistem dikatakan berhasil bila mengenai sasaran atau tujuan yang telah direncanakan.

2.1.3. Klasifiksi Sistem

Sistem merupakan suatu bentuk integrasi antara satu komponen dengan komponen lain karena sistem memiliki sasaran yang berbeda untuk setiap kasus yang terjadi yang ada di dalam sistem tersebut. Oleh karena itu, sistem dapat diklasifikasikan dari beberapa sudut pandang di antaranya[5. p, 22] :

a. Sistem abstrak dan sistem fisik

Sistem abstrak adalah sistem yang berupa pemikiran atau ide-ide yang tidak tampak secara fisik, misalnya sistem teologia, yaitu sistem yang berupa pemikiran hubungan antara manusia dengan Tuhan, sedangkan sistem fisik merupakan sistem yang ada secara fisik, misalnya sistem komputer, sistem produksi, sistem penjualan, sistem administrasi personalia dan lain sebagainya.

b. Sistem alamiah dan sistem buatan manusia

Sistem alamiah adalah sistem yang terjadi melalui proses alam; tidak dibuat oleh manusia, misalnya sistem perputaran bumi, terjadinya siang malam, pergantian musim. Sedangkan sistem buatan manusia merupakan sistem yang melibatkan interaksi manusia dengan mesin yang disebut human machine sistem. Sistem informasi berbasis komputer merupakan contoh human machine sistem karena menyangkut penggunaan komputer yang berinteraksi dengan manusia.

c. Sistem determinasi dan sistem probobalistik

Sistem yang beroperasi dengan tingkah laku yang dapat diprediksi disebut sistem deterministic. Sistem komputer adalah contoh dari sistem yang tingkah lakunya dapat dipastikan berdasarkan program-program komputer yang dijalankan, sedangkan sistem yang bersifat probabilistik adalah sistem yang kondisi masa depannya tidak dapat diprediksi karena mengandung unsur probablistic.

d. Sistem terbuka dan sistem tertutup

Sistem tertutup merupakan sistem yang tidak berhubungan dan tidak terpengaruh oleh lingkungan luarnya. Sistem ini bekerja secara otomatis tanpa campur tangan pihak luar. Sedangkan sistem terbuka adalah sistem yang berhubungan oleh lingkungan luarnya. Sistem ini menerima masukan dan menghasilkan keluaran untuk subsistem lainnya.

Menurut uraian diatas dapat disimpulkan bahwa sebuah sistem memiliki beberapa komponen yang di klasifikasikan berdasarkan sudut pandang yang bertujuan untuk menentukan tujuan tertentu.

2.2. Konsep Dasar Informasi

2.2.1. Pengertian Informasi

Informasi merupakan sesuatu yang sangat berpengaruh di dalam kehidupan manusia , dalam dalam sebuah organisasi. Informasi sendiri dapat diartikan sebagai berikut :

Informasi adalah data yang diolah menjadi bentuk yang berguna bagi pemakainya.[2. p,36]

Informasi adalah data yang diolah menjadi bentuk yang lebih berguna bagi penerimanya[6. p,8]

Berdasarkan pengertian di atas maka dapat disimpulkan bahwa pengertian informasi adalah suatu data yang telah di olah sedemikian rupa sehingga dapat bermanfaat bagi pemaikai untuk pengambilan keputusan.

2.2.2. Siklus Informasi

Siklus informasi adalah gambaran secara umum mengenai proses terhadap data sehingga menjadi informasi yang bermanfaat bagi pengguna. Informasi yang menghasilkan informasi berikutnya. Demikian seterusnya proses pengolahan data menjadi informasi.

Data merupakan bentuk mentah yang belum dapat bercerita banyak, sehingga perlu diolah lebih lanjut. Data ditangkap sebagai input, diproses melalui suatu model membentuk informasi. Pemakai kemudian menerima informasi tersebut sebagai landasan untuk membuat suatu keputusan dan melakukan tindakan operasional yang akan membuat sejumlah data baru. Data baru tersebut selanjutnya menjadi input pada proses berikutnya, begitu seterusnya sehingga membentuk suatu siklus informasi/*Information Cycle*.[8. p,17]

2.2.3. Kualitas Informasi

Kualitas informasi sangat dipengaruhi atau ditentukan oleh tiga hal pokok, yaitu akurasi (*accuracy*), relevansi (*relevancy*), dan tepat waktu (*timeliness*).[7. p,274]

a. Akurasi (*accuracy*)

Sebuah informasi harus akurat karena dari sumber informasi hingga penerima informasi kemungkinan banyak terjadi gangguan yang dapat mengubah atau merusak informasi tersebut. Informasi dikatakan akurat apabila informasi tersebut tidak bias atau menyesatkan, bebas dari kesalahan-kesalahan dan harus jelas mencerminkan maksudnya.

Ketidakakuratan sebuah informasi dapat terjadi karena sumber informasi (data) mengalami gangguan atau kesengajaan sehingga merusak atau mengubah data-data asli tersebut.

Beberapa hal yang dapat berpengaruh terhadap keakuratan sebuah informasi antara lain adalah:

1. Informasi yang akurat harus memiliki kelengkapan yang baik, karena bila informasi yang dihasilkan sebagian tentunya akan memengaruhi dalam pengambilan keputusan atau menentukan tindakan secara keseluruhan, sehingga akan berpengaruh terhadap kemampuannya untuk mengontrol atau memecahkan suatu masalah dengan baik.
2. Informasi yang dihasilkan oleh proses pengolahan data, haruslah benar sesuai dengan perhitungan-perhitungan yang ada dalam proses tersebut.

3. Informasi harus aman dari segala gangguan (noise) dapat mengubah atau merusak akurasi informasi tersebut dengan tujuan utama.

b. Tepat waktu (*timeliness*)

Informasi yang dihasilkan dari suatu proses pengolahan data, datangnya tidak boleh terlambat (usang). Informasi yang terlambat tidak akan mempunyai nilai yang baik, karena informasi merupakan landasan dalam pengambilan keputusan. Kesalahan dalam mengambil keputusan akan berakibat fatal bagi perusahaan. Mahalnya informasi disebabkan harus cepat dan tepat informasi tersebut didapat. Hal itu disebabkan oleh kecepatan untuk mendapatkan, mengolah dan mengirimkan informasi tersebut memerlukan bantuan teknologi-teknologi terbaru. Dengan demikian diperlukan teknologi-teknologi mutakhir untuk mendapatkan, mengolah, dan mengirimkan informasi tersebut.

c. Relevansi (*relevancy*)

Informasi dikatakan berkualitas jika relevan bagi pemakainya. Hal ini berarti bahwa informasi tersebut harus bermanfaat bagi pemakainya. Relevansi informasi untuk tiap-tiap orang satu dengan lainnya berbeda. Misalnya, informasi mengenai kerusakan infrastruktur laboratorium komputer ditujukan kepada rektor universitas. Tetapi akan lebih relevan apabila ditujukan kepada penanggung jawab laboratorium

2.3. Sistem Informasi

2.3.1. Definisi Sistem Informasi

Berikut ini adalahh beberapa pengerian dari sistem informasi menurut beberapa ahli :

Sistem informasi adalah suatu komponen yang terdiri dari manusia, teknologi informasi, dan prosedur kerja yang memproses, menyimpan, menganalisis, dan menyebarkan informasi untuk mencapai suatu tujuan.[7. p,29]

Sistem informasi merupakan suatu sistem yang tujuannya menghasilkan informasi.[9. p,33]

Berdasarkan dari pengertian sistem informasi menurut para ahli yang telah dipaparkan di atas dapat ditarik kesimpulan bahwa sistem informasi adalah gabungan dari komponen – komponen seperti manusia , teknologi informasi , dan prosedur kerja yang bertujuan untuk menghasilkan suatu informasi yang dapat berguna untuk mencapai tujuan tertentu.

2.3.2. Komponen Sistem Informasi

Sistem informasi terdiri dari komponen-komponen yang disebut dengan istilah blok bangunan (building block). Sebagai suatu sistem, blok bangunan tersebut masing-masing berinteraksi satu dengan yang lainnya membentuk satu kesatuan untuk mencapai sasarannya. Blok bangunan tersebut terdidiri dari : [9. p,47]

1. Blok Masukan (*Input Block*)

Input mewakili data yang masuk ke dalam sistem informasi. Input disini termasuk metode-metode dan media untuk menangkap data yang akan dimasukkan, yang dapat berupa dokumen-dokumen dasar.

2. Blok Model (*Model Block*)

Blok ini terdiri dari kombinasi prosedur, logika dan model matematik yang akan memanipulasi data input dan data yang tersimpan di basis data dengan cara yang sudah tertentu untuk menghasilkan keluaran yang diinginkan.

3. Blok Keluaran (*Output Block*)

Produk dari sistem informasi adalah keluaran yang merupakan informasi yang berkualitas dan dokumentasi yang berguna untuk semua tingkatan manajemen serta semua pemakai sistem

4. Blok Teknologi (*Technology Block*)

Teknologi digunakan untuk menerima input, menjalankan model, menyimpan dan mengakses data, menghasilkan dan mengirimkan keluaran dan membantu pengendalian sistem secara keseluruhan. Pada blok ini terdiri dari 3 bagian utama, yaitu teknisi (*humanware* atau *brainware*), perangkat lunak (*software*) dan perangkat keras (*hardware*).

5. Blok Basis Data (*Database Block*)

Basis data (*database*) merupakan kumpulan dari data yang saling berhubungan satu dengan yang lainnya, tersimpan di perangkat keras komputer dan digunakan perangkat lunak untuk memanipulasinya. Basis data diakses atau dimanipulasi dengan menggunakan perangkat lunak paket yang disebut dengan DBMS (*Database Management Systems*).

6. Blok Kendali (*Controls Block*)

Beberapa pengendalian perlu dirancang dan diterapkan untuk meyakinkan bahwa hal-hal yang dapat merusak sistem dapat dicegah ataupun bila terlanjur terjadi kesalahan-kesalahan dapat langsung cepat diatasi.

2.4. Plagiat / *Plagiarism*

2.4.1. Pengertian Plagiat

Plagiat adalah perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyebutkan sumber secara tepat dan memadai.[10. p,9]

2.4.2. Ruang Lingkup Plagiat

Ketentuan dalam pasal 2 PERMENDIKNAS No.17, juga menjelaskan ruang lingkup plagiat bahwa tindakan yang termasuk dalam plagiat meliputi 5 tindakan namun tidak terbatas pada : [11. p,9]

1. Mengacu dan/atau mengutip istilah, kata-kata dan/atau kalimat, data dan/atau informasi dari suatu sumber tanpa menyebutkan sumber dalam catatan kutipan dan/atau tanpa menyatakan sumber secara memadai;
2. Mengacau dan/atau mengutip secara acak istilah, kata-kata dan/atau kalimat, data dan/atau informasi dari suatu sumber tanpa menyebutkan sumber dalam catatan kutipan dan/atau tanpa menyatakan sumber secara memadai;

3. Menggunakan sumber gagasan, pendapat, pandangan, atau teori tanpa menyatakan sumber secara memadai
4. Merumuskan dengan kata-kata dan/atau kalimat sendiri dari sumber kata-kata dan/atau kalimat, gagasan, pendapat, pandangan, atau teori tanpa menyatakan sumber secara memadai.
5. Menyerahkan suatu karya ilmiah yang dihasilkan dan/atau telah dipublikasikan oleh pihak lain sebagai karya ilmianya tanpa menyatakan sumber secara memadai.

2.5. Konsep Basis Data

Berikut ini adalah beberapa pengertian dari basis data yang dipaparkan menurut para ahli :

Basis data adalah salah satu bagian dari sistem informasi secara keseluruhan. [12. p,24]

Basis : Gudang , markas, tempat bersarang atau berkumpul.

Data :Representasi fakta dunia nyata yang mewakili suatu objek seperti : manusia(pegawai, siswa, pembeli, pelanggan), barang, hewan, peristiwa, konsep, keadaan dan sebagainya yang direkam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi atau kombinasinya.

Secara sederhana database (basis data) dapat diungkapkan sebagai suatu pengorganisasian data dengan bantuan komputer yang memungkinkan data dapat diakses dengan mudah dan cepat.[13. p,2]

2.6. Internet

Internet merupakan jaringan komputer yang dibentuk oleh Departemen Pertahanan Amerika Serikat pada tahun 1969, melalui proyek lembaga ARPA yang mengembangkan jaringan yang dinamakan ARPANET (Advanced Research Project Agency Network), di mana mereka mendemonstrasikan bagaimana dengan hardware dan software komputer yang berbasis UNIX.

Tujuan awal dibangunnya proyek itu adalah untuk keperluan militer. Pada saat itu Departemen Pertahanan Amerika Serikat (US Department of Defense) membuat sistem jaringan komputer yang tersebar dengan menghubungkan komputer di daerah-daerah vital untuk mengatasi masalah bila terjadi serangan nuklir dan untuk menghindari terjadinya informasi terpusat, yang apabila terjadi perang dapat mudah dihancurkan.

Pada mulanya ARPANET hanya menghubungkan 4 situs saja yaitu Stanford Research Institute, University of California, Santa Barbara, University of Utah, di mana mereka membentuk satu jaringan terpadu pada tahun 1969, dan secara umum ARPANET diperkenalkan pada bulan Oktober 1972. Tidak lama kemudian proyek ini berkembang pesat di seluruh daerah, dan semua universitas di negara tersebut ingin bergabung, sehingga membuat ARPANET kesulitan untuk mengaturnya.

Oleh sebab itu ARPANET dipecah manjadi dua, yaitu "MILNET" untuk keperluan militer dan "ARPANET" baru yang lebih kecil untuk keperluan non-

militer seperti, universitas-universitas. Gabungan kedua jaringan akhirnya dikenal dengan nama DARPA Internet, yang kemudian disederhanakan menjadi Internet.

2.6.1. Pengertian Internet

Internet kependekan dari *interconnection-networking*, ialah sistem global dari seluruh jaringan komputer yang saling terhubung menggunakan standar *Internet Protocol Suite* (TCP/IP) untuk melayani miliaran pengguna di seluruh dunia.Internet adalah jaringan komputer dunia yang menghubungkan jaringan-jaringan komputer regional diseluruh dunia. [14]

2.7. Perangkat Lunak Pendukung

2.7.1. XAMPP

XAMPP adalah sebuah software yang berfungsi untuk menjalankan website berbasis PHP dan menggunakan pengolah data MYSQL di komputer lokal". XAMPP berperan sebagai server web pada komputer lokal. XAMPP juga dapat disebut sebuah Cpanel server virtual, yang dapat membantu melakukan preview sehingga dapat dimodifikasi website tanpa harus online atau terakses dengan internet. [15. p,7]

2.7.2. MySql

MySQL adalah salah satu database manajemen sistem (DBMS) dari sekian banyak DBMS seperti Oracle, MS SQL, Postgre SQL, dan lainnya". Dimana MySQL dalam operasi client-server melibatkan server daemon MySQL disisi server dan berbagai macam program serta library yang berjalan besar. SQL singkatan dari Structure Query Language dan sering disebut Sequel saja. SQL mulai

dikembangkan tahun 70-an di laboratorium IBM, Stan Jose, California. Untuk mengakses sebuah file database, salah satu server database yang kecil dan mudah digunakan namun memiliki kehandalan dan performa tinggi.[16. p,45]

2.7.3. PHP (*Hypertext Preprocessor*)

Hypertext Preprocessor (PHP) yaitu bahasa pemograman web server-side yang bersifat open source". PHP merupakan script yang terintegrasi dengan HTML dan berada pada server (server side HTML embedded scripting). PHP adalah script yang digunakan untuk membuat halaman website yang dinamis. Dinamis berarti halaman yang akan ditampilkan dibuat saat halaman itu diminta oleh client. Mekanisme ini menyebabkan informasi yang diterima client selalu yang terbaru. Semua script PHP dieksekusi pada server di mana script tersebut dijalankan. [16. p,3]

2.7.4. HTML (*Hypertext Markup Language*)

Hypertext Markup Language (HTML) adalah sekumpulan simbol-simbol atau tag-tag yang dituliskan dalam sebuah file yang digunakan untuk menampilkan halaman pada web browser. [16. p,40]

2.7.5. CSS (*Cascading Style Sheet)*

CSS merupakan singkatan dari *Cascading Style Sheet*. Kegunaannya adalah untuk mengatur tampilan dokumen HTML, contohnya seperti pengaturan jarak antar baris, teks, warna dan format border bahkan penampilan file gambar."

CSS dikembangkan oleh W3C. organisasi yang mengembangkan teknologi internet. Tujuannya tak lain untuk mempermudah proses penataan halaman web. [17.p,2]

2.7.6. JavaScript

JavaScript adalah bahasa *scripting* yang populer di internet dan dapar bekerja di sebagian besar browser populer seperti internet explorer , mozila firefox , netscape , dan opera. Kode javascript dapat disisipkan dalam halaman web menggunakan tag script.[18.p,17]

2.8. Algoritma Rabin-Karp

Algoritma Rabin Karp ditemukan oleh Michael O. Rabin dan Richard M. Karp pada tahun 1987 yang menggunakan fungsi hashing untuk menemukan pattern didalam string teks. Fungsi hashing menyediakan metode sederhana untuk menghindari perbandingan jumlah karakter yang quadratik di dalam banyak kasus atau situasi. Dari pada melakukan pemeriksaan terhadap setiap posisi dari teks ketika terjadi pencocokan pola, akan lebih efisien untuk melakukan pemeriksaan hanya jika teks yang sedang proses memiliki kemiripan seperti pada pattern. Untuk melakukan pengecekan kemiripan antara dua kata ini digunakan fungsi hash.[19. p,5]

2.8.1 *Hashing*

Hashing adalah suatu cara untuk mentransformasikan sebuah *string* menjadi suatu nilai yang unik dengan panjang tertentu (*fixed-length*) yang berfungsi

sebagai penanda *string* tersebut. Fungsi untuk menghasilkan nilai ini disebut fungsi *hash*, sedangkan nilai yang dihasilkan disebut hash. [19. p,5]

2.8.2. *K-Gram*

K-grams adalah rankaian term dengan panjang *k*. Kebanyakan yang digunakan sebagai terms adalah kata. *K-grams* merupakan sebuah metode yang diaplikasikan untuk pembangkitan kata atau karakter. Metode *K-grams* ini digunakan untuk mengambil potongan-potongan karakter huruf sejumlah *k* dari sebuah kata yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen. [19. p,5]

2.8.3. *Rolling Hash*

Algoritma Rabin-Karp menggunakan rolling hash untuk mencari *string* yang sama. *Rolling hash* melakukan perhitungan *hash* dengan cara mengambil *substring* dalam suatu *string* sebanyak k karakter. *String* inputan akan dikelompokkan menjadi *substring* dengan panjang m karakter kemudian dihitung *hash*-nya berdasar nilai ASCII dari tiap karakter, rumus menghitung rolling hash adalah: [19. p,5]

$$H = c_1 a^{k-1} + c_2 a^{k-2} + c_3 a^{k-3} + ... + c_k a^0$$

Gambar 2.1 Rumus menghitung *Rolling Hash*

(sumber http://www.slideshare.net/mukaninawi/sistem-pengukuran-tingkat-similaritas-dokumen-dengan-algoritma-rabinkarp-dan-enhanced-confix-stripping-stemmer [18]*)*

Keterangan:

H = nilai hash, *c* = nilai ASCII suatu karakter.

a = basis (tidak boleh 1 dan 0).

k = banyaknya karakter.

untuk pola yang panjang dan teks yang besar, algoritma ini menggunakan operasi *mod*, setelah dikenai operasi *mod q*, nilai akan menjadi lebih kecil dari nilai *q*.

Tetapi tidak semua nilai hash yang sama polanya juga sama. Hal ini sering terjadi pada beberapa kasus, ini disebut *spurious hits*. Kemungkinan terjadinya diantaranya karena: [19. p,5]

a. Operasi *mod* terinterfensi oleh keunikan nilai *hash* (nilai *mod q* biasanya dipilih bilangan prima sehingga 10*q* hanya cocok dengan 1 kata komputer).

b. Informasi hilang setelah penjumlahan. [18. p,5]

2.8.4 *Dice's Similiarity Coefficient*

Untuk menghitung kemiripan dapat digunakan *Dice Similarity Coeficients* dengan cara menghitung jumlah K-Gram yang digunakan pada kedua dokumen yang diuji. Nilai kemiripan tersebut dapat dihitung dengan rumus :

$$s = \frac{2n_t}{n_x + n_y}$$

Gambar 2.2 Rumus *Dice Similarity Coeficients*

(sumber:https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient#Formula)

dimana S adalah nilai kemiripan, n_t adalah jumlah K-Gram yang sama dan n_x serta n_y masing-masing adalah jumlah K-Gram dari masing-masing string yang diujikan. Hasil yang didapat adalah bilangan dari 0 sampai 100 dimana bilangan tersebut adalah persentase kemiripan 2 string yang diujikan.