

PROSIDING **SEMINAR** **KEBANGSAAN** **SAINS, TEKNOLOGI & SAINS SOSIAL**

27 ~ 28 MEI 2002

HOTEL VISTANA, KUANTAN, PAHANG

Anjuran :



Universiti Teknologi MARA
Cawangan Pahang

Dengan Kerjasama



Kerajaan
Negeri Pahang Darul Makmur

JILID 1



PERKEMBANGAN TEORI MAKLUMAT/KOMUNIKASI

BAHARI BIN IDRUS¹, MOHAMMAD ALINOR BIN ABDUL KADIR² & TRINI PHILIPS²

¹Jabatan Komputeran Industri, Fakulti Teknologi & Sains Maklumat, UKM, 43600 Bangi, Selangor.

²Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi, UKM, 43600 Bangi, Selangor.

ABSTRAK

Memandangkan pesatnya pembangunan sains maklumat/komunikasi yang dikecapi oleh Malaysia dewasa ini, maka perlulah kita untuk melihat serba sedikit tentang bagaimana teori-teori maklumat/komunikasi silam dimodelkan daripada fahaman-fahaman falsafah yang menyelubunginya. Untuk itu, penulisan ini berhasrat untuk membuat sorotan kepada perkembangan teori maklumat/komunikasi semenjak awal kewujudannya.

Katakunci: *Teori Komunikasi, Claude E. Shannon, Permodelan Sumber Maklumat, Ergodik*

PENGENALAN

Perkataan Inggeris "communication" berasal daripada perkataan Latin *communico* bermaksud untuk berkongsi. Perkataan "communication" juga berasal daripada kata dasar "common" yang berkait dengan perbuatan untuk membuat sesuatu (contohnya maklumat) supaya diketahui ramai. Perkataan dasar "common" juga membawa kepada perkataan "community". Proses untuk memberitahu dan membuatkan maklumat diketahui umum mengimplicasikan bahawa terdapat suatu hubungan di antara pemberitahu dan yang diberitahu. Sebarang bentuk komunikasi antara manusia mengandungi satu ciri sistem yang disambung menjadi rangkaian. Satu sistem komunikasi boleh merupakan dawai telefon, udara atau saraf optik manusia. Sistem-sistem mestilah berpasangan antara satu sama lain semata-mata untuk memindahkan maklumat dan keadaan sebarang sistem yang menyatukan. Sekiranya alat penghubung ini terpisah, maklumat tidak dapat dipindahkan (Severin & Tankard, 1988).

Proses fizikal yang menghasilkan, mengekalkan, menyalurkan dan penghasilan semula isyarat merupakan teori matematik komunikasi yang juga dikenali sebagai teori maklumat atau teori saluran isyarat. Teori yang termoden dihasilkan oleh Claude Shannon(1948) semasa beliau menjadi seorang penyelidik matematik di Makmal Telefon Bell dan juga seorang Profesor Sains di Institut Teknologi Massachusetts. Penulisan beliau yang bertajuk "*The Mathematical Theory of Communication*" telah menjadi rangsangan tunggal paling penting untuk perkembangan model-model dan teori-teori lain dalam komunikasi. Teori ini telah menyediakan pencetus untuk banyak lagi model dalam teori komunikasi.

SEJARAH AWAL TEORI KOMUNIKASI

Menurut Cherry (1978), pada tahun 1267, Roger Bacon telah mencadangkan "a certain sympathetic needle (lodestone)" boleh digunakan untuk komunikasi jarak jauh. Pada abad ke-16, Porta dan Gilbert telah menulis tentang "sympathetic telegraph" dan pada tahun 1746, Watson di England telah menghantar isyarat elektrik melalui wayar sepanjang lebih kurang 2 meter. Pada tahun 1753, seorang pekerja yang tidak dikenali telah menggunakan satu pasang wayar untuk setiap abjad. Kemudian, pada tahun 1787, Lomard telah menggunakan satu pasang wayar dan beberapa simbol. Pengenalan kepada "gelombang pembawa" semasa Perang Dunia Pertama telah dipraktikkan dengan penciptaan penapisan gelombang oleh G.A Campbell. Peruntukan isyarat yang serentak kepada frekuensi jalur gelombang telah menjadi asas untuk komunikasi secara elektrik. Teknik yang telah mendorong perkembangan teori komunikasi secara umum adalah melalui penciptaan telefon dan televisyen. Telefon dicipta oleh Alexander Graham Bell pada tahun 1876. "Hingar" merupakan satu batas utama kepada komunikasi. Ia merujuk kepada sebarang gangguan selain daripada isyarat yang diperlukan atau mesej yang dipilih dan dihantar. Namun begitu, teori komunikasi sebenarnya berasal daripada kajian tentang komunikasi elektrik dan beberapa idea penting tentang teori komunikasi ini akan diutarakan di sini.

TEORI KOMUNIKASI ABAD KE-19

Pada tahun 1832, Samuel F.B Morse telah memulakan kajian tentang telegraf elektrik. Telegraf ini melukis garisan yang panjang dan pendek pada jalur kertas dan jujukan ini diwakili dengan nombor yang ditandakan dengan perkataan atau dikenali sebagai kod Morse pada tahun 1937. Ini merupakan cara pengekodan yang agak efisien tetapi agak menyusahkan. Semasa Morse bekerja dengan Alfred Vail, pengekodan cara lama ini ditinggalkan dan diganti dengan kod Morse yang dicipta pada tahun 1838. Menurut Pierce (1961), dalam kod ini perkataan dan abjad diwakili dengan ruang, titik dan tanda sempang. Ruang bermaksud ketiadaan elektrik, titik menandakan arus elektrik pada masa yang singkat dan tanda sempang merupakan arus elektrik dalam pada masa yang panjang. Pada amnya, gabungan pendek titik dan tanda sempang digunakan untuk huruf yang selalu digunakan dan gabungan panjang digunakan untuk huruf yang jarang digunakan. Dalam telegraf arus tunggal, terdapat dua elemen yang boleh digunakan untuk membina kod: arus dan tiada arus yang boleh dipanggil 1 atau 0. Dalam telegraf berganda, terdapat tiga elemen: arus terkehadapan yang merupakan arus ke dalam wayar; tiada arus; arus kebelakang iaitu arus keluar dari wayar; atau boleh juga ditandakan sebagai +1, 0, -1. Di sini + atau - menandakan arah arus elektrik dan nombor 1 menandakan magnitud kekuatan arus.

Pada tahun 1874, Thomas Edison memperkenalkan sistem telegraf ganda empat (kuadrupleks). Empat keadaan arus elektrik yang berbeza dengan dua mesej dihantar melalui satu litar boleh diwakili sebagai +3, +1, -1, -3. Pada tahun 1855, William Thomson telah mengira secara tepat apakah arus yang akan diterima apabila titik atau tanda sempang dihantar melalui kabel kapal selam. Satu pendekatan terhadap masalah-masalah tersebut telah dibuat melalui penciptaan telefon oleh Alexander Graham Bell. Penggunaan telefon tidak menggunakan isyarat off-on yang perlahan seperti yang digunakan dalam telegraf sebaliknya menggunakan arus yang kekuatannya berubah-ubah pada kadar amplitud yang besar dengan kepantasan beratus kali ganda berbanding dengan menggunakan telegraf secara manual.

Beberapa orang lain yang turut menyumbang teori matematik bagi telefon adalah Henri Poincare yang merupakan seorang ahli matematik Perancis, G.A Campbell dari Syarikat Telefon dan Telegraf Amerika dan Joseph Fourier. Kajian Fourier tentang pengaliran haba telah diaplikasikan kepada kajian tentang getaran dan merupakan alat yang penting untuk menganalisis sifat arus elektrik yang berubah-ubah mengikut masa dalam bentuk yang agak kompleks seperti arus elektrik yang dihasilkan melalui penggunaan telefon dan telegraf.

TEORI KOMUNIKASI NYQUIST

Henry Nyquist merupakan ahli matematik yang telah menyelesaikan masalah telegraf. Beliau menyelesaikan masalah telegraf dengan kaedah yang mantap dan menerbitkan satu kertas ilmiah bertajuk "*Certain Factors Affecting Telegraph Speed*" pada tahun 1924. Pierce (1961) menyatakan bahawa kertas kerja ini menjelaskan hubungan antara kelajuan telegraf dan bilangan nilai arus elektrik seperti +1, -1 (dua nilai arus elektrik) atau +3, +1, -1, -3 (empat nilai arus elektrik). Nyquist menyatakan bahawa jika kita menghantar simbol pada kadar yang tetap, di sini ia merupakan nilai arus yang berjaya dihantar, kelajuan transmisi W adalah berkait dengan m bilangan simbol m yang berlainan atau nilai arus elektrik yang wujud, iaitu $W = K \log m$, dengan K adalah malar dan nilainya bergantung kepada berapa banyak nilai arus elektrik yang berjaya dihantar pada setiap saat. Logaritma adalah fungsi yang sebaiknya dalam hubungan Nyquist. Andaikan dikhususkan dua pilihan tak bersandar "off" atau "on", 0 atau 1, secara serentak. Terdapat empat kemungkinan gabungan dua pilihan tak bersandar pilihan 0 atau 1, seperti yang ditunjukkan dalam jadual di bawah:

Jadual 1: Empat kemungkinan gabungan dua pilihan tak bersandar

Bilangan gabungan	Pilihan Pertama 0 atau 1	Pilihan Kedua 0 atau 1
1	0	0
2	0	1
3	1	0
4	1	1

Seterusnya untuk menentukan tiga pilihan tak bersandar 0 atau 1 pada waktu yang sama, lapan gabungan akan didapati.

Jadual 2: Lapan kemungkinan gabungan tiga pilihan tak bersandar

Bilangan gabungan	Pilihan Pertama 0 atau-1	Pilihan Kedua 0 atau 1	Pilihan Ketiga 0 atau 1
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Jika ingin menentukan empat pilihan tak bersandar 0 atau 1, maka didapati enam belas gabungan yang berbeza.

Jika dapat ditentukan M gabungan tak bersandar 0 atau 1 secara serentak, maka M mesej tak bersandar serentak dapat dihantar, dan pastilah kelajuan bersandar kepada M. Tetapi untuk menghantar M mesej secara serentak, didapati terdapat 2^M gabungan yang mungkin daripada M pilihan tak bersandar 0 atau 1. Oleh itu untuk menghantar mesej secara serentak, diperlukan 2^M simbol atau nilai arus yang berbeza. Andaikan 2^M simbol yang berbeza dipilih. Nyquist menyatakan logaritma kepada bilangan simbol diambil supaya mendapat kelajuan garisan dan $\log 2^M = M$. Oleh yang demikian bilangan simbol hanyalah merupakan bilangan pilihan tak bersandar 0 atau 1 yang boleh diwakili secara serentak iaitu bilangan mesej yang tak bersandar secara serentak.

Hubungan Nyquist menyatakan bahawa dengan beralih daripada telegraf “off-on” kepada tiga arus (+1, 0, -1) telegraf kelajuan menghantar huruf atau simbol yang lain ditingkatkan kepada 60 peratus. Manakala jika menggunakan empat arus (+3,+1, -1, -1) kelajuan dapat digandakan. Seterusnya Nyquist menunjukkan penggunaan lapan nilai arus (0, 1, 2, 3, 4, 5, 6, 7, atau +7, +5, +3, +1, -1, -3, -5, -7) seharusnya membolehkan penghantaran empat kali lebih laju berbanding dengan dua nilai arus. Tetapi beliau sedar bahawa keadaan yang berubah-ubah dalam pengecilan arus, gangguan atau hingar dan batas terhadap kuasa yang boleh digunakan menjadikan penggunaan nilai arus yang banyak agak sukar.

Beralih kepada masalah kadar untuk elemen isyarat yang boleh dihantar, Nyquist telah mentakrifkan kelajuan garis sebagai $1 \frac{1}{2}$ daripada bilangan elemen isyarat (titik, ruang, nilai arus) yang boleh dihantar dalam masa satu saat Nyquist menerbitkan pada tahun 1928 kertas kerja kedua beliau yang bertajuk “*Certain Topics in Telegraph Transimission Theory.*” Dalam kertas kerja ini beliau menunjukkan bahawa jika seseorang menghantar sebilangan $2N$ nilai arus yang berbeza setiap saat, semua komponen sinusoidal isyarat dengan frekuensi lebih tinggi daripada N adalah berulang dan jika semua frekuensi yang tinggi ini dikeluarkan, seseorang masih lagi boleh membuat kesimpulan dengan mengkaji isyarat dimana nilai arus dihantar.

TEORI KOMUNIKASI HARTLEY

R.V.L. Hartley, pencipta osiloskop meringkaskan ideanya dalam sebuah kertas kerja “*Transmission of Information*”, yang diterbitkan pada tahun 1928. Hartley mempunyai cara yang menarik untuk menformulasikan masalah komunikasi. Beliau mengandaikan penghantar mesej disediakan dengan satu set simbol yang secara mental dia memilih simbol demi simbol sehingga membentuk jujukan simbol. Beliau memerhatikan bahawa peristiwa peluang adalah seperti memutar bola ke dalam poket yang mungkin akan menjanakan jujukan seumpama di atas. Beliau kemudiannya mentakrifkan maklumat H untuk mesej sebagai logaritma kepada bilangan jujukan simbol yang mungkin dipilih dan menunjukkan bahawa $H = n \log s$, dengan n merupakan bilangan simbol yang dipilih dan s merupakan bilangan simbol yang berbeza dalam set yang mana simbol dipilih.

Dalam kes ini logaritma s, iaitu bilangan simbol, merupakan bilangan pilihan 0 atau 1 yang tak bersandar yang boleh dihantar secara serentak dan adalah wajar kadar transmisi maklumat merupakan kadar penghantaran simbol per saat n didarabkan dengan bilangan pilihan 0 atau 1 yang tak bersandar yang boleh dibawa setiap simbol. Hartley juga telah mendekati masalah mengekod simbol utama (contohnya huruf untuk abjad) dalam istilah simbol sekunder (contohnya jujukan titik dalam kod Morse). Beliau memerhatikan bahawa sekatan kepada pilihan simbol (fakta bahawa E dipilih lebih kerap daripada Z) haruslah mengawal panjang simbol sekunder jika hendak menghantar simbol dengan lebih berkesan. Morse sendiri memahami perkara ini tetapi Hartley telah menyatakan perkara ini dalam sedemikian cara yang telah mendorong pendekatan secara matematik dan menginspirasi kajian seterusnya. Hartley juga telah mencadangkan cara untuk mengaplikasikan perkara di atas kepada isyarat yang berterusan seperti isyarat telefon atau isyarat gambar.

Akhirnya Hartley menyatakan, seperti juga Nyquist, bahawa jumlah maklumat yang boleh dihantar adalah berkadar dengan lebar jalur didarabkan dengan masa penghantaran.. Selepas kajian yang dilakukan oleh Nyquist dan Hartley ternyata teori komunikasi kekal seperti sebelumnya tanpa ada perkembangan yang menarik. Tetapi keadaan menjadi lebih rumit dan kompleks semasa perang Dunia Kedua. Banyak pemahaman tentang sistem komunikasi khusus yang baru dicapai tetapi tidak ada dasar falsafah yang luas dihasilkan. (Pierce 1961)

PEMODELAN SUMBER MAKLUMAT SHANNON

Claude E. Shannon(1948) menerbitkan artikel bertajuk “*The Mathematical Theory of Communication*” di Jurnal Bell System Technical pada Julai dan Oktober 1948. Ideanya telah mencetuskan satu sensasi dan telah berkembang secara luas dalam dua perkembangan yang utama: iaitu teori maklumat yang menggunakan kebarangkalian dan teori ergodik untuk mengkaji ciri-ciri statistik data dan sistem komunikasi dan teori pengekodan yang kebanyakannya menggunakan kebarangkalian dan geometri untuk membina kod yang efisien untuk pelbagai situasi. Dalam proses mengusahakan teori komunikasi yang boleh diaplikasikan dan membina sistem telekomunikasi yang lebih baik, beliau telah mentakrifkan ukuran entropi ($H = \sum p_i \log p_i$) iaitu apabila diaplikasikan kepada sumber maklumat, ia boleh menentukan kapasiti saluran yang diperlukan untuk menghantar sumber itu sebagai digit perduaan yang telah dikodkan. Bahagian seterusnya di dalam kertas ini hanya akan membincangkan aspek pemodelan sumber maklumat dan sumber ergodik tersebut.

Komunikasi manusia mungkin boleh di terangkan dengan menggunakan model rawak atau model statistik. Sebagai contohnya, dalam mengira kekerapan penggunaan huruf E dalam prosa Inggeris, didapati 0.13 daripada huruf yang muncul adalah E sementara W merangkumi 0.02 daripada semua huruf yang muncul. Tetapi, didapati kadar E dan W yang hampir sama akan ditemui dalam prosa yang ditulis oleh sesiapa sahaja. Oleh itu, kita boleh membuat ramalan secara lebih yakin bahawa anda, saya atau sesiapa sahaja apabila menulis teks yang panjang, artikel ataupun buku dalam bahasa Inggeris, sekurang-kurangnya 0.13 daripada huruf yang ditulis adalah E.

Shannon(1948) telah mendemonstrasikan bagaimana perkataan dan teks Inggeris boleh dianggarkan dengan menggunakan proses matematik yang boleh dilakukan oleh sebuah mesin. Andaikan sebagai contoh jujukan huruf dan ruang dengan kebarangkalian yang sama dihasilkan. Ini boleh dilakukan dengan memasukkan bilangan kad yang sama yang telah ditanda dengan setiap huruf dan ruang ke dalam sebuah kotak, mencampuradukkan dan mengambil satu kad, merekod simbol, mengembalikannya, mencampuradukkannya sekali lagi, mengambil satu kad dan lakukan perkara yang sama. Ini akan memberikan apa yang dipanggil oleh Shannon sebagai penghampiran sifar bagi bahasa Inggeris.

1) Penghampiran sifar:

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSHYD QPAAMKBZAACIBZLHJQD

Perhatikan terdapat terlalu banyak Z dan W dan tidak cukup E dan ruang. Penghampiran kepada teks Inggeris boleh dilakukan dengan memilih huruf secara tak bersandar tapi memilih huruf E lebih kerap berbanding W atau Z. Ini boleh dilakukan dengan memasukkan lebih banyak huruf E dan mengurangkan pemasukan W dan Z ke dalam kotak, mencampuradukkannya dan mengeluarkan huruf-huruf tersebut. Kebarangkalian yang diberi kepada E haruslah 0.013 iaitu setiap seratus huruf yang dimasukkan ke dalam kotak, 13 daripadanya mestilah huruf E. Kebarangkalian sesuatu huruf itu adalah W mestilah 0.02 iaitu

daripada seratus huruf dimasukkan ke dalam kotak, 2 mestilah huruf W dan seterusnya. Ini akan memberikan apa yang dipanggil oleh Shannon sebagai penghampiran pertama Inggeris.

2) Penghampiran pertama:

OCRO HLI RGWR NWIELWIS EU LL NBNESBYA TH EEI ALHENHTTPA OOBTTVA
NAH BRL

Dalam teks Inggeris, kemungkinan untuk berhadapan dengan pasangan huruf yang bermula Q adalah tidak mungkin kecuali QU. Kebarangkalian untuk berhadapan dengan QX atau QZ adalah sifar. Sementara kebarangkalian QU tidak sifar, tetapi ia adalah sangat kecil sehingga tidak disenaraikan dalam jadual. Sementara itu, kebarangkalian TH adalah 0.037, kebarangkalian WE adalah 0.0006. Kebarangkalian-kebarangkalian ini membawa maksud seperti berikut. Katakanlah sebuah teks mengandungi 10,001 huruf. Terdapat 10,000 pasangan huruf iaitu huruf yang pertama dengan yang kedua, huruf kedua dengan yang ketiga dan seterusnya sehinggalah kepada huruf yang kedua terakhir dengan yang terakhir. Daripada pasangan-pasangan tersebut, terdapat sebilangannya adalah pasangan huruf TH. Kemungkinan terdapat 370 pasang huruf ini. Jika jumlah bilangan TH dijumpai yang diandaikan 370 kali dibahagikan dengan jumlah bilangan pasangan huruf yang diandaikan 10,000 akan didapati kebarangkalian bahawa pasangan huruf yang dipilih secara rawak dalam teks tersebut ialah TH iaitu $370/10,000$ atau 0.037.

Ahli kriptografi telah membina jadual seperti kebarangkalian digram untuk teks Inggeris. Kebarangkalian digram bermaksud kebarangkalian huruf tertentu akan diikuti oleh huruf tertentu yang lain. Teks Inggeris boleh dibina dengan menggunakan kebarangkalian digram. Ini boleh didemonstrasikan seperti berikut. Andaikan digram yang pelbagai dimasukkan dalam beberapa kotak. Satu digram akan diambil secara rawak daripada mana-mana kotak dan huruf ini akan direkodkan. Kemudian digram yang kedua akan dikeluarkan daripada kotak yang ditandakan oleh huruf kedua daripada digram pertama dan huruf kedua daripada digram kedua ini dicatatkan. Seterusnya digram ketiga akan diambil daripada kotak yang ditandakan oleh huruf kedua daripada digram kedua dan huruf kedua digram ketiga tersebut dicatatkan. Proses sebegini diteruskan. Ruang akan dianggap sama seperti huruf. Terdapat kebarangkalian khusus bahawa ruang akan diikuti oleh huruf tertentu dan kebarangkalian khusus bahawa huruf tertentu diikuti oleh ruang. Dengan menggunakan proses yang sama, Shannon telah membina penghampiran kedua terhadap teks Inggeris iaitu:

3) Penghampiran kedua:

ON IE ANSTOUTINYS ARE T INCTROE ST BE S DEAMY ACHIN D ILONASIVE
TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

Ahli kriptografi juga telah membina jadual yang memberi kebarangkalian kepada sekumpulan tiga huruf yang dipanggil kebarangkalian trigram. Ini membolehkan Shannon untuk membina penghampiran ketiga untuk teks Inggeris iaitu:

4) Penghampiran ketiga:

IN NO I ST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF
DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

Persamaan terhadap teks Inggeris dapat dilihat meningkat langkah demi langkah melalui contoh 1 hingga 4 yang diberi oleh Shannon. Contoh 1 iaitu penghampiran sifar, tidak terdapat struktur yang boleh difahami dan adalah mustahil untuk mengenal pasti jujukan ini apabila ia datang dari sebarang bahasa tertentu yang menggunakan abjad yang sama. Dalam contoh 2, walaupun jujukan yang ditunjukkan di sini tidak layak sebagai bahasa Inggeris yang bagus, tetapi ia ada mempamerkan beberapa struktur bahasa Inggeris. Seterusnya, perhatikan dalam contoh (3), bagaimana ciri-ciri bahasa Inggeris daripada jujukan ini dirasai daripada penghampiran yang kedua. Kita tidak akan menghadapi banyak masalah untuk mengenal pasti jujukan ini sebagai penghampiran bahasa Inggeris dan bukannya bahasa Perancis, misalnya dalam contoh 4 yang mengambilkira frekuensi trigram, terdapat lapan perkataan Inggeris.

Jelas kelihatan daripada contoh-contoh yang diberi bahawa dengan memberikan mesin statistik bahasa yang tertentu dan dengan memberi mesin itu keupayaan yang menyerupai mengambil bola daripada kotak, membaling duit syiling atau memilih nombor secara rawak, mesin tersebut dapat menghasilkan penghampiran yang dekat dengan teks Inggeris atau teks dalam bahasa lain. Daripada segi struktur dan statistiknya mengikut pandangan manusia, lebih lengkap maklumat yang diberi kepada mesin semakin hampirlah produknya menyerupai Inggeris atau teks lain. Oleh itu adalah lebih mudah jika mesin dibekalkan dengan perkataan berbanding dengan huruf dan membiarkannya untuk menghasilkan perkataan

ini mengikut kebarangkaliannya. Shannon telah memberikan contoh dimana perkataan dipilih secara tak bersandar tetapi dengan kebarangkalian ia untuk berlaku dalam teks Inggeris. Ini boleh dilakukan dengan menggantung teks kepada perkataan, mencampuraduknya dalam kotak dan seterusnya mengambil perkataan secara berturut-turut. Beliau memanggilnya sebagai penghampiran pertama perkataan. Ia adalah seperti berikut:

- 5) Penghampiran Pertama Perkataan
**REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN
DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT
GRAY COME TO FURNISHE THE LINE MESSAGE HAD TO BE THESE**

Beliau memilih pasangan perkataan yang pertama secara rawak dalam sebuah novel. Seterusnya helaian muka surat dibelek sehinggalah berjumpa dengan huruf kedua pasangan pertama perkataan. Seterusnya perkataan yang menyusuli huruf kedua pasangan pertama perkataan itu akan dicatitkan dan seterusnya proses ini akan diteruskan. Proses ini telah memberikan beliau penghampiran kedua perkataan Inggeris.

- 6) Penghampiran Kedua Perkataan
**THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISHWRITER THAT THE
CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE
LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN
UNEXPECTED**

Dapat dilihat bahawa terdapat banyak perkataan dalam pernyataan ini yang menggambarkan dan sesungguhnya mungkin berlaku dalam teks Inggeris. Pertimbangkan apakah sebenarnya yang ingin disampaikan di sini. Dalam teks Inggeris yang sebenar contohnya teks yang dihantar oleh mesin teletaip, huruf tertentu berlaku dengan frekuensi yang hampir tetap. Pasangan huruf dan huruf ganda tiga dan ganda empat juga berlaku dengan frekuensi yang hampir tetap. Perkataan dan pasangan perkataan juga berlaku dengan frekuensi yang hampir tetap. Seterusnya Shannon (1948) menyatakan bahawa dengan menggunakan proses matematik secara rawak yang dilakukan oleh sebuah mesin boleh menghasilkan jujukan perkataan atau huruf Inggeris yang mempamerkan statistik sebegini.

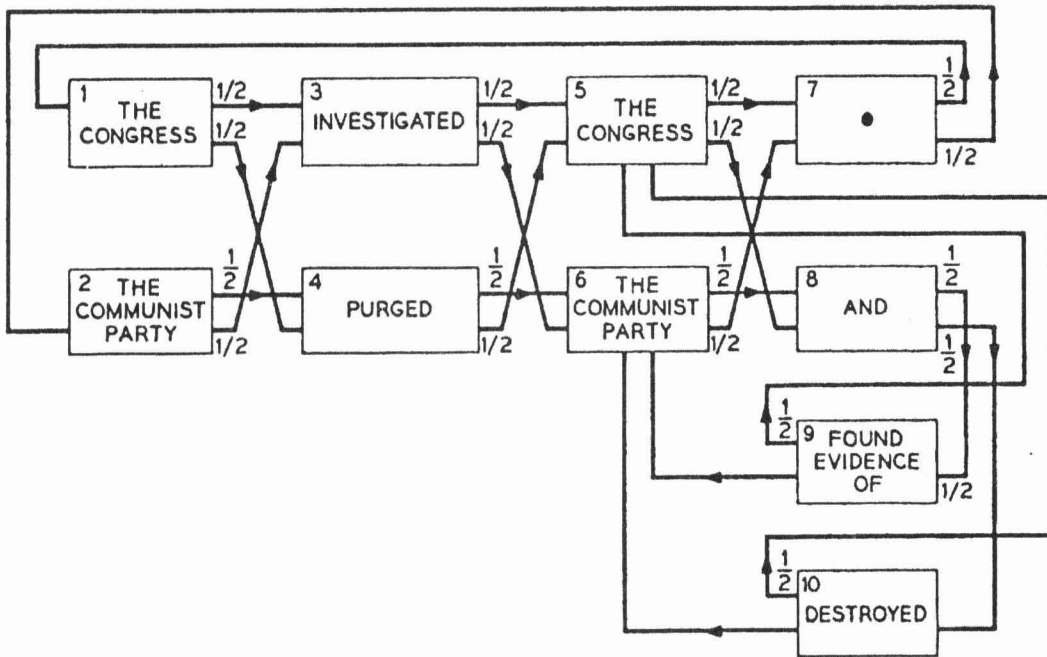
Rajah 1 merupakan diagram “mesin” yang dimaksudkan. Pierce (1961) telah memberikan penerangan tentang mesin tersebut. Setiap kotak yang dinomborkan mewakili keadaan mesin tersebut. Oleh kerana hanya terdapat bilangan kotak atau keadaan yang terhingga, maka ia dipanggil sebagai mesin keadaan terhingga. Beberapa anak panah dilukis dari sebuah kotak ke sebuah kotak yang lain. Dalam mesin yang khusus ini, hanya dua anak panah akan pergi dari sebuah kotak ke sebuah kotak yang lain. Dalam kes ini juga setiap anak panah dilabelkan dengan $\frac{1}{2}$. Ini menunjukkan bahawa kebarangkalian mesin akan melalui contohnya dari keadaan 2 ke keadaan 3 adalah $\frac{1}{2}$ dan kebarangkalian mesin itu melalui dari keadaan 2 ke keadaan 4 adalah $\frac{1}{2}$. Untuk membolehkan mesin ini beroperasi, jujukan pilihan rawak boleh dilakukan dengan melabung duit syiling secara berturutan. Biarkan kepala (H) bermaksud anak panah di sebelah atas diikuti. Dan ekor (T) bermaksud anak panah di sebelah bawah diikuti. Ini menunjukkan keadaan yang baru yang akan dilalui. Apabila ini dilakukan, perkataan tersebut akan dicetak ataupun simbol yang ditulis dalam kotak dan duit syiling akan dilambung sekali lagi untuk mendapat keadaan yang baru.

Sebagai contoh, jika bermula di keadaan 7 dan jujukan kepala dan ekor seperti dilambung berikut: T H H H T T H T T T H H H H, mesin itu akan mencetak jujukan perkataan seperti berikut:

**THE COMMUNIST PARTY INVESTIGATED THE CONGRESS. THE COMMUNIST PARTY
PURGED THE CONGRESS AND DESTROYED THE COMMUNIST PARTY AND FOUND
EVIDENCE OF THE CONGRESS.**

Pilihan secara rawak mengikut jadual kebarangkalian jujukan simbol atau perkataan boleh menghasilkan bahan yang menyerupai teks Inggeris. Keadaan terhingga mesin dengan pilihan rawak membolehkan peralihan daripada satu keadaan ke satu keadaan boleh menghasilkan bahan yang menyerupai teks Inggeris. Kedua-dua proses ini dipanggil proses stokastik kerana elemen rawak yang terlibat di dalamnya.

Rajah 1: Mesin yang dimaksudkan oleh Shannon. (Pierce, 1961)



Sebenarnya melalui penerangan serta contoh-contoh yang diberi memperjelaskan tentang model matematik untuk sumber teks. Model sebegini haruslah berkeupayaan untuk menghasilkan teks yang hampir menyerupai teks Inggeris yang sebenar, sangat hampir sehinggakan masalah mengekod dan menghantar teks sebegini adalah pada dasarnya setara dengan masalah mengekod dan menghantar teks Inggeris yang sebenar. Sifat-sifat berkenaan matematik model tersebut haruslah ditakrifkan secara matematik supaya teorem yang berguna berkenaan pengekodan dan penghantaran teks boleh dihasilkan. Model secara matematik yang diadaptasikan oleh Shannon untuk mewakili penghasilan teks merupakan sumber ergodik. Maka penerangan tentang sumber ergodik akan diperjelaskan seterusnya. Menarik untuk dikaji seterusnya, apakah model Shannon ini masih berlangsung untuk Bahasa Melayu, samada menggunakan huruf Latin atau huruf Jawi.

SUMBER ERGODIK

Bagi ahli matematik dan statistik matematik, konsep ergodik dan keadaan di mana sesuatu punca itu dikatakan ergodik adalah mendalam. Walaubagaimanapun, untuk matlamat kajian ilmiah ini, konsep sumber ergodik yang termudah saja akan digunakan. Menurut Abramson (1990), suatu sumber ergodik hanya merupakan sumber yang mana jika dibuat cerapan untuk suatu jangka masa yang panjang, akan (dengan kebarangkalian 1) mengeluarkan jujukan simbol punca yang “tipikal”.

Sebenarnya idea tentang sumber yang memiliki ciri-ciri ergodik adalah terlalu lazim sehinggakan sukar untuk membayangkan sumber yang tak ergodik. Oleh itu sumber tak ergodik diberikan.

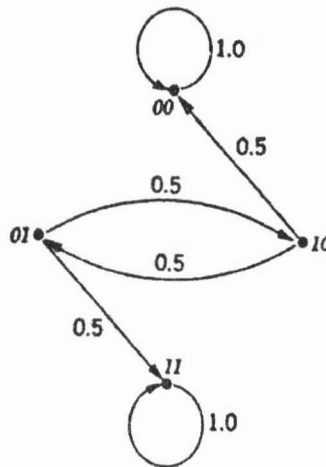
Contoh: Katakan diberi punca perduaan $S \{0,1\}$. Diandaikan kebarangkalian simbol bersyarat

$$P(0/00) = P(1/11) = 1.0$$

$$P(1/00) = P(0/11) = 0$$

$$P(0/01) = P(0/10) = P(1/01) = P(1/10) = 0.5$$

Terdapat empat keadaan –00,01,10,11. Gambarajah keadaan bagi punca ini ditunjukkan dalam Rajah 2
 Rajah 2: Keadaan bagi suatu sumber tak ergodik



Perhatikan bahawa bagi sumber ini, jika kita berada sama ada di keadaan 00 atau 11, kita akan kekal di keadaan ini selama-lamanya. Seterusnya, andaikan kita memilih salah satu daripada empat keadaan yang ada secara rawak (iaitu, setiap keadaan mempunyai kebarangkalian $\frac{1}{4}$ untuk dipilih). Dari itu, jika kita mulai dalam keadaan yang telah terpilih, kita tahu bahawa selepas banyak kali alihan keadaan berlaku kita akan berada di keadaan 00 dengan kebarangkalian 0.5. Iaitu, selepas begitu banyak simbol dikeluarkan daripada sumber, sumber ini akan mengeluarkan 0 dengan kebarangkalian 0.5 dan 1 dengan kebarangkalian 0.5. Walau bagaimanapun, jika kita tunggu lebih lama lagi, kita pasti akan melihat di dalam sebarang jujukan tertentu yang dikeluarkan oleh sumber ini akan terdiri daripada sama ada semuanya sifar atau semuanya satu. Dengan perkataan lain (dengan kebarangkalian 1) kita tidak akan melihat jujukan yang tipikal daripada sumber ini iaitu ia tidak ergodik

Jenis punca stokastik yang dipilih sebagai model sumber mesej adalah sumber ergodik. Sumber ergodik merupakan kebarangkalian atau sumber mesej stokastik yang ringkas dan pendekatan secara matematik untuk proses yang ringkas adalah lebih mudah. Model matematik yang diketengahkan adalah memadai untuk mewakili beberapa aspek tentang peranan manusia sebagai sumber mesej dan memadai untuk mewakili beberapa aspek tentang mesej yang dihasilkan. Sebagai contoh, dengan mengambil teks Inggeris, didapati frekuensi kejadian huruf yang tertentu berlaku secara tetap. Didapati juga jujukan huruf dengan frekuensi yang serupa dengan teks Inggeris boleh dilakukan dengan menggunakan proses rawak atau stokastik seperti menggantung teks kepada huruf atau perkataan, mencampuradukkannya dalam kotak dan mengambilnya satu per satu. Proses stokastik yang lebih rumit dengan menggunakan mesin keadaan terhingga boleh menghasilkan anggaran teks Inggeris yang lebih tepat. Teorem-teorem teori maklumat yang diperbincangkan oleh Shannon diaplikasikan terhadap sumber ergodik dan bukti-buktinya adalah berdasarkan kepada andaian bahawa sumber mesej itu adalah ergodik.

RUJUKAN

1. Cherry, C. 1978. *On human communication*. Ed. Ke-3. United States of America: Prentice-Hall, Inc.
2. Severin, W.J. & Tankard, J.W. 1998. *Teori Komunikasi*. Terj. Lowe, V. & Abdullah bin Hassan. Kuala Lumpur. Dewan Bahasa dan Pustaka.
3. Pierce, J.R. 1961. *Symbols, signals and noise*. United States of America: Harper & Row Publishers.
4. Shannon, C.E. 1948. A Mathematical Theory of Communication. Dlm. Slepian, D. *Key papers in the development of Information Theory*, hlm. 5-18. United States of America: The Institute of Electrical and Electronics Engineers, Inc
5. Abramson, N. 1990. *Teori Maklumat dan pengkodan*. Terj. Khairuddin Ab Hamid. Kuala Lumpur: Dewan Bahasa dan Pustaka.