

Text (Web) Mining dan Pemanfaatan FOSS (Free and Open Source Software) Pendukungnya

Dr. Taufik Fuadi Abidin, S.Si., M.Tech

Ketua Program Studi Informatika
FMIPA Universitas Syiah Kuala

tfa@informatika.unsyiah.ac.id
www.informatika.unsyiah.ac.id/tfa

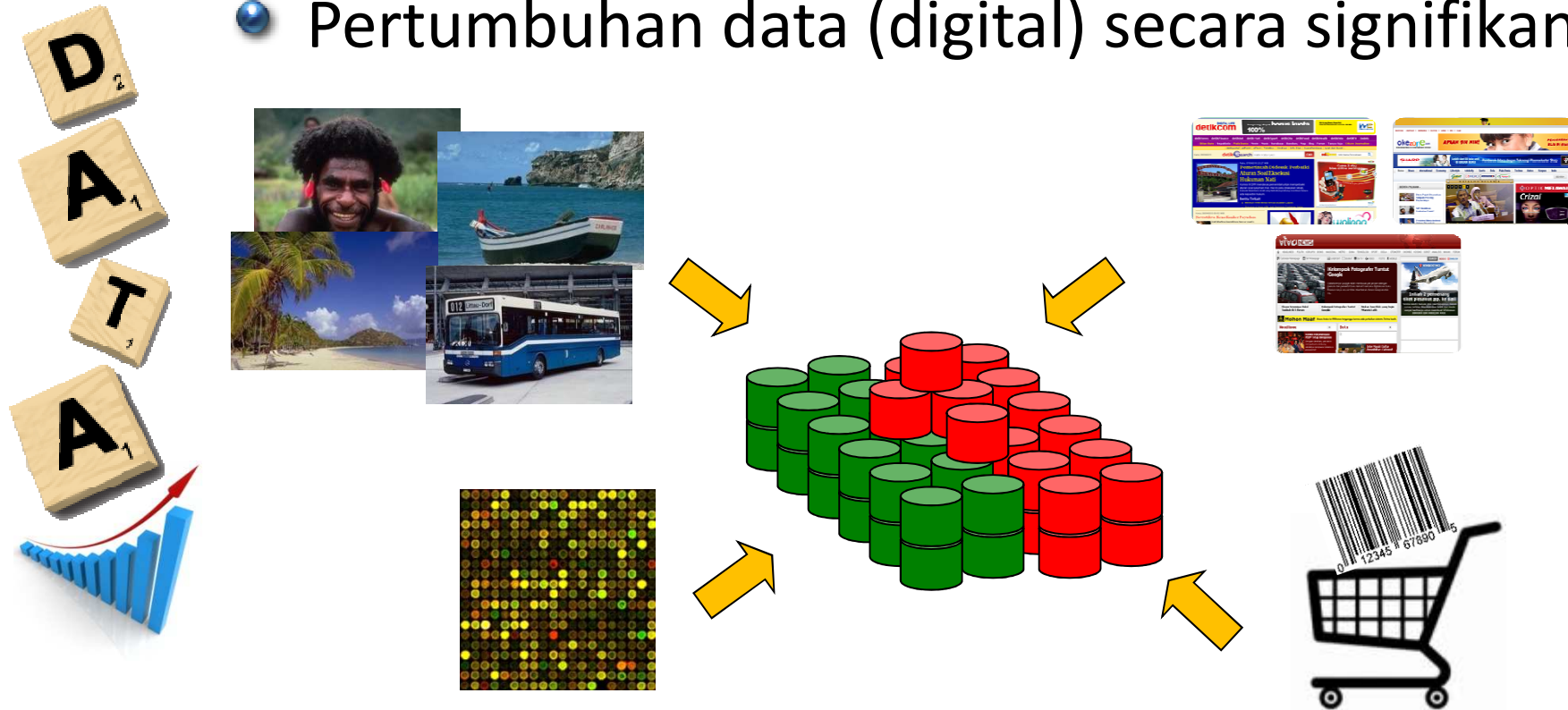
Disampaikan pada

Seminar Nasional Teknologi Informasi dan Komunikasi (SNASTIKOM) 2012

Medan, 31 Januari 2012

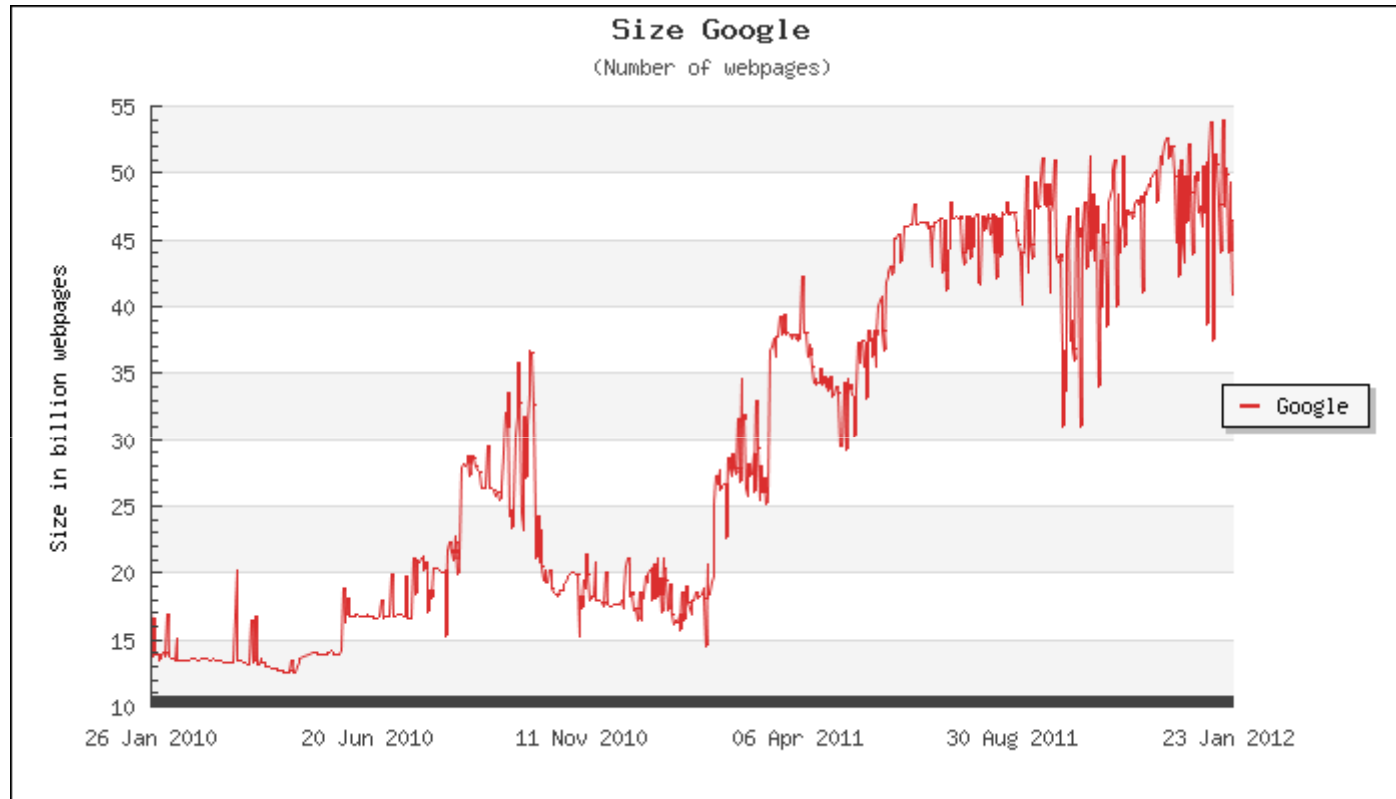
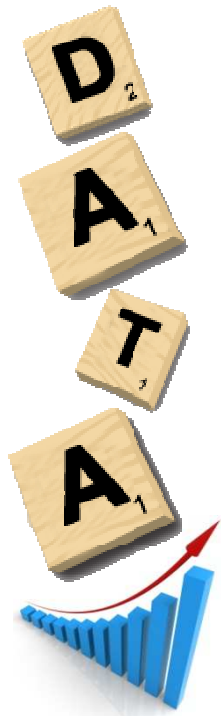
Mengapa Data Mining?

- Pertumbuhan data (digital) secara signifikan



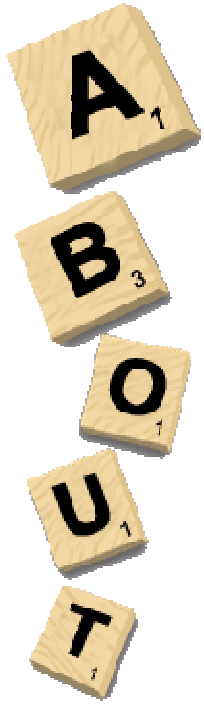
- Data melimpah tidak akan bermakna apapun bila tidak diolah dan dianalisa

Ukuran Web (Estimasi)



Source: http://www.worldwidewebsize.com/Graph_v1.php?Searchengine=Google&corpus=0&Last=730

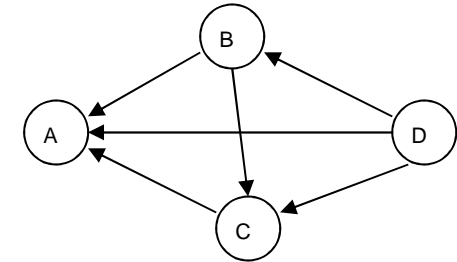
Text/Web Mining



- Mencari pola (*pattern*) yang menarik dan bermanfaat dari kumpulan data teks/web (*text/web database*) dalam jumlah besar menggunakan komputer
- Data teks atau web adalah tidak (semi) terstruktur, data relational terstruktur
- Penelitian tentang teks dan web mining akan terus berkembang karena jumlah data teks dan web terus meningkat



Web Mining



• Web Structure Mining

- Adanya hyperlink dan interkoneksi antar hyperlink dapat memberikan informasi menarik, contoh: Pagerank (Google)
- *Text scoring* berdasarkan kemunculan kata (*text apperance*) dalam **Anchor Tag dan Title**

• Web Content Mining

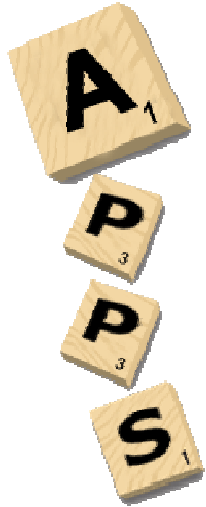
- Klasifikasi halaman web berdasarkan topik
- Pengelompokan (*clustering*) halaman web

• Web Usage Mining

- Pattern dari logs (*users access based on clicks*)



Aplikasi Text/Web Mining



- *Learning to Classify Tropical Disease Web Pages from Large Indonesian Web Documents*

The 4th International Conference on Computer and Electrical Engineering (ICCEE 2011), Singapore October 14-15, 2011

- *Penentuan Secara Otomatis Akronim dan Ekspansinya Dari Data Teks Berbahasa Indonesia*

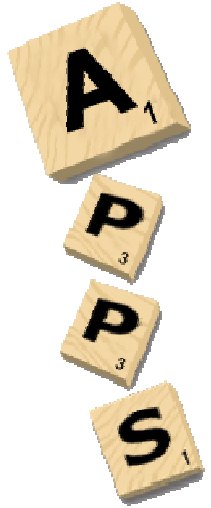
Seminar Nasional Teknologi Informasi, Komunikasi dan Aplikasinya, Klub Bunga Butik Resort, Batu, Malang 11 Nov 2011

- *Affix Stemming untuk Bahasa Indonesia*

Seminar Nasional dan Expo Teknik Elektro, Banda Aceh, 24 Oktober 2011



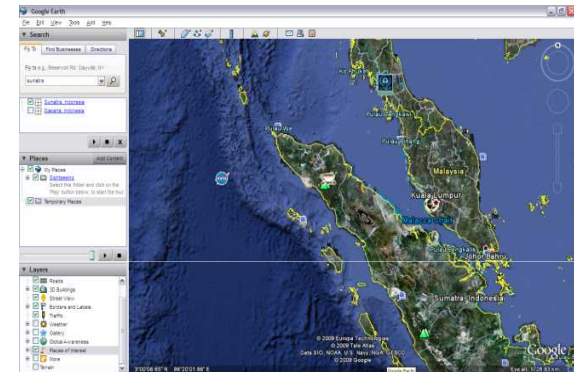
Web Classification



- *Learning to Classify Tropical Disease Web Pages from Large Indonesian Web Documents*



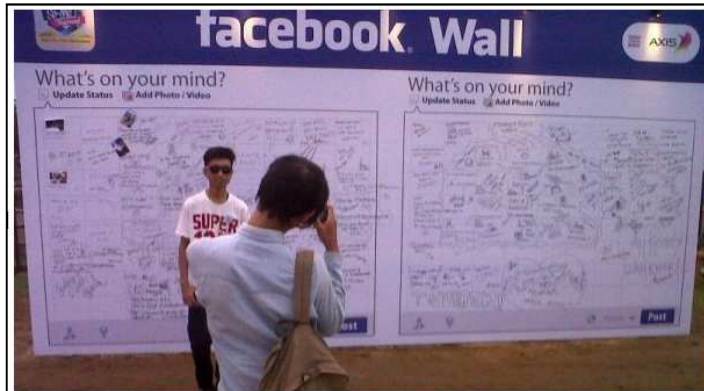
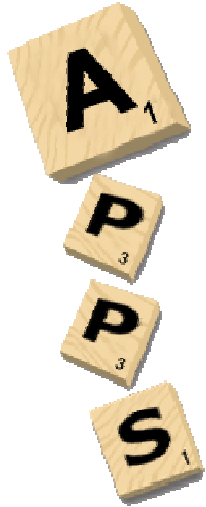
Extract Spatio-Temporal Data of Tropical Disease Cases



How can we do this?



Fakta: Topik Web Sangat Variatif



Facebook Wall (Foto: Yoga/okezone)

CALIFORNIA - Dalam sebuah survei terungkap bahwa rata-rata pengguna Facebook menghabiskan 32 menit setiap hari membaca update di jejaring sosial dari teman-teman dan keluarga. Selanjutnya, lebih dari setengah atau 56 persen rata-rata orang menulis status update kira-kira dua baris panjang.

Survei yang dilakukan s
Facebook rajin memper

KOMPAS C E T A K

/ Home

Penyakit Tropis Tidak Teratasi

Jumlah Penderita Tak Kunjung Turun

Senin, 11 Agustus 2008 | 00:56 WIB

Jakarta, Kompas - Penyebaran sejumlah penyakit tropis, seperti malaria, kusta atau lepra, dan filariasis atau kaki gajah, hingga kini tidak teratasi. Hal ini ditandai dengan angka kasus yang dalam lima tahun terakhir tidak kunjung turun.

Penyakit kaki gajah, misalnya, pada tahun 2000 berjumlah 4.472 penderita, tahun 2005 berdasarkan data Departemen Kesehatan melonjak menjadi 10.239 penderita.

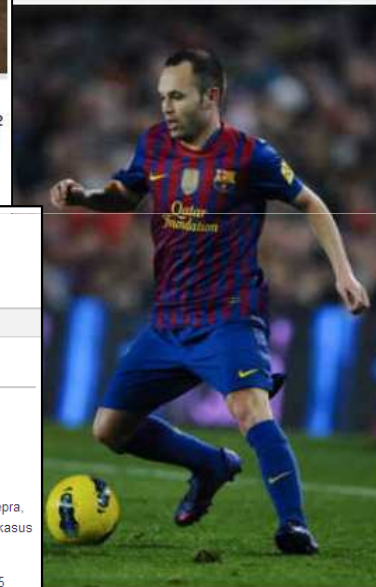
Di beberapa daerah, kasus penyakit kaki gajah juga mengalami lonjakan. Di Jawa Barat, misalnya, tahun 2000 kasus kaki gajah hanya ditemukan di Kota dan Kabupaten Bekasi.

"Sekarang, sampai Juni 2008 tercatat 413 kasus di seluruh 26 kabupaten dan kota di Jabar serta tersebar di 240 desa dan 133 kecamatan," kata Kepala Subdinas Penyehatan



19/09/08 16:27

10 Kasus Flu Burung di Aceh Hingga September 2008



Getty Images/David Ramos

Barcelona - Kepastian soal kondisi Andres Iniesta akhirnya didapat. Cedera yang didapat gelandang Barcelona itu akan memaksanya absen bermain selama tiga pekan. Sementara Alexis Sanchez akan absen dua pekan.

Iniesta gagal menuntaskan laga El Clasico di ajang Copa del Rey, Kamis (26/1/2012) dinihari WIB tadi. Dia terpaksa ditarik keluar di menit 30 dan digantikan Pedro Rodrigues.

Sebelumnya dilaporkan kalau pesepakbola 29 tahun itu mengalami perobekan di otot kaki kirinya. Dan setelah dilakukan pemeriksaan lebih mendetil terhadap kondisinya, Iniesta

akan absen sampai tiga pekan ke depan. Demikian diberitakan

10 kasus flu burung

elmi di Banda Aceh, Langsa (dua kasus)

ni tidak terdapat

ahkan karena

oto serta 41

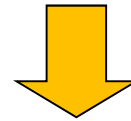
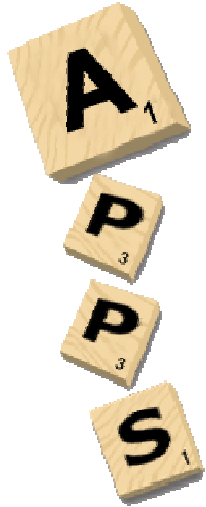


INFORMATIKA

www.informatika.unsyiah.ac.id



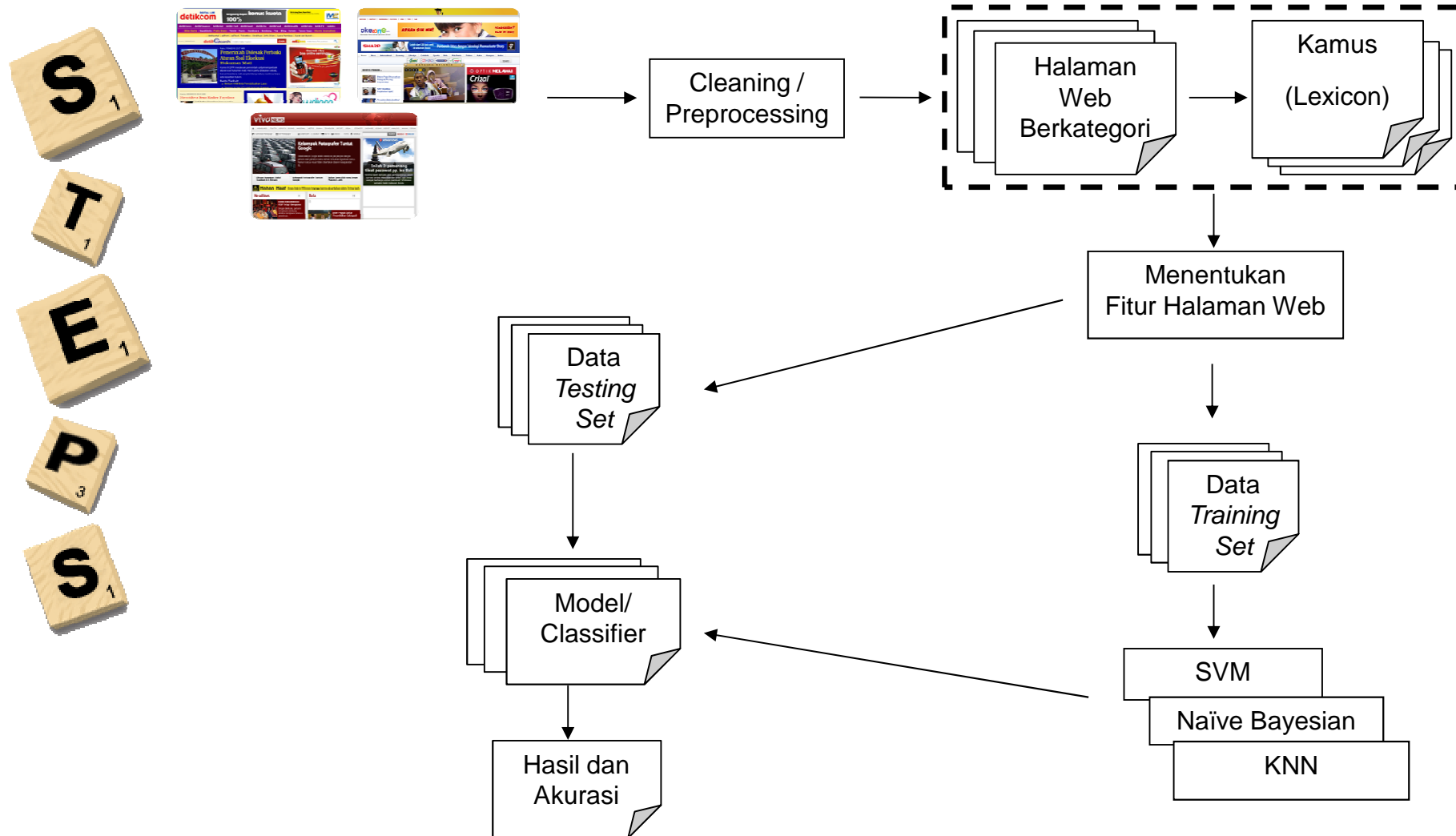
Web Classifier is Required!



Web Classifier



Steps: Overview



Preprocessing dan FOSS Tool

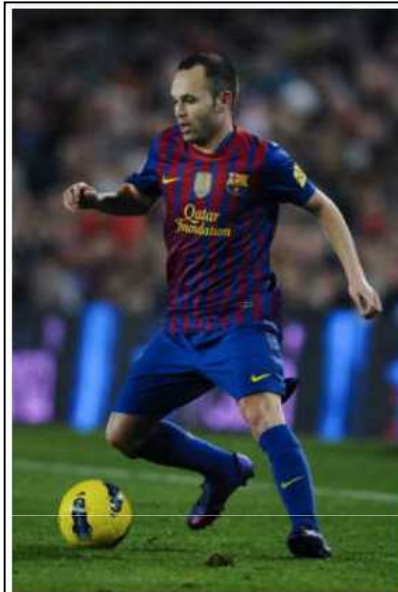


FOTO:Getty Images/David Ramos

dipastikan akan absen sampai tiga pekan ke depan. Demikian diberitakan Reuters.

Barcelona - Kepastian soal kondisi Andres Iniesta akhirnya didapat. Cedera yang didapat gelandang Barcelona itu akan memaksanya absen bermain selama tiga pekan. Sementara Alexis Sanchez akan absen dua pekan.

Iniesta gagal menuntaskan laga *El Clasico* di ajang Copa del Rey, Kamis (26/1/2012) dinihari WIB tadi. Dia terpaksa ditarik keluar di menit 30 dan digantikan Pedro Rodriguez.

Sebelumnya dilaporkan kalau pesepakbola 29 tahun itu mengalami perobekan di otot kaki kirinya. Dan setelah dilakukan pemeriksaan lebih mendetil terhadap kondisinya, Iniesta



<http://www.perl.org/>



CPAN

The Comprehensive Perl Archive Network (CPAN) has over 18,000 open source modules available for download.

<http://www.cpan.org/>

HTML-Extract-Content

Module HTML-ExtractContent-0.10.tar.gz

Module Exporter-Lite-0.02.tar.gz

Module Class-Accessor-Lvalue-0.11.tar.gz

Module Class-Accessor-0.34.tar.gz

Module Want-0.18.tar.gz

```
<br>Iniesta gagal menuntaskan laga <a  
href="http://www.detiksport.com/sepakbola/read/2012/01/26/045850/1825177/75/ditahan-  
madrid-2-2-barcelona-lolos-ke-  
semifinal"><strong><em>ElClasico</em></strong>  
></a> di ajang Copa del Rey, Kamis  
(26/1/2012) dinihari WIB tadi.
```



INFORMATIKA

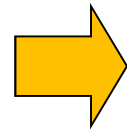
www.informatika.unsyiah.ac.id



Preprocessing: Remove HTML Tags!



<http://www.detiksport.com/sepakbola/read/2012/01/26/231628/1826540/75/iniesta-menepi-tiga-pekan-alexis-dua>



HTML-Extract-Content



File Edit View Search Terminal Help

```
<title>sepak bola iniesta menepi tiga pekan alexis dua</title>
<content>barcelona kepastian soal kondisi andres iniesta akhirnya didapat. c
edera yang didapat gelandang barcelona itu akan memaksanya absen bermain sel
ama tiga pekan. sementara alexis sanchez akan absen dua pekan. iniesta gagal
menuntaskan laga el clasico di ajang copa del rey kamis 26/1/2012 dinihari wi
b tadi. dia terpaksa ditarik keluar di menit 30 dan digantikan pedro rodrigu
es. sebelumnya dilaporkan kalau pesepakbola 29 tahun itu mengalami perobekan
di otot kaki kirinya. dan setelah dilakukan pemeriksaan lebih mendetil terh
adap kondisinya iniesta dipastikan akan absen sampai tiga pekan ke depan. de
mikian diberitakan reuters. itu berarti iniesta tak akan bisa memperkuat the
catalans dalam laga semifinal copa del rey menghadapi levante atau valencia
yang akan digelar pekan depan. namun pesepakbola internasional spanyol itu
akan diharapkan akan pulih saat barca menjalani leg pertama babak 16 besar l
iga champions kontra bayer leverkusen di 14 februari. kabar lain terkait pem
ain barca adalah terkait kondisi alexis sanchez. mengalami cedera dalam laga
yang sama usai bertabrakan dengan pepe pemain depan asal chile itu terpaksa
menepi dua pekan. kehilangan dua pemain tersebut bisa jadi akan mempengaruh
i kekuatan barca dalam upayanya mengejar real madrid di puncak klasemen. jos
ep guardiola sebelumnya sudah kehilangan david villa ibrahim afellay dan and
reu fontas yang didera cedera. sementara seydou keita sedang membela mali di
piala afrika dan maxwell akan hengkang ke paris st germain di akhir bulan i
ni. din roz</content>
```

1,1

All

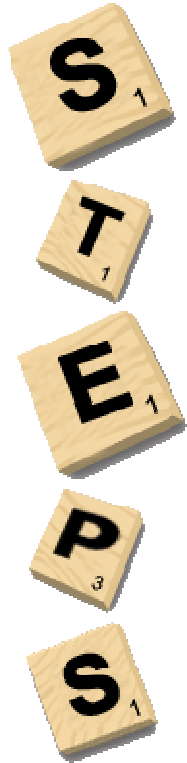


INFORMATIKA

www.informatika.unsyiah.ac.id



Langkah Selanjutnya



- Kamus 1-gram, 2-grams dan 3-grams untuk kelas positif dan negatif dibangun
- Fitur tekstual dari setiap halaman web (data pembelajaran) dibangkitkan dari bagian judul (*title*), konten-atas, konten-tengah, dan konten-bawah
- Rasio dari setiap kata 1-gram, 2-grams, dan 3-grams pada keempat bagian tersebut dihitung untuk setiap klas (*class*)
- Jumlah fitur = 24, yang diperoleh dari 3 jenis grams x 4 bagian pengamatan x 2 klas

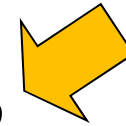


Top 2-grams



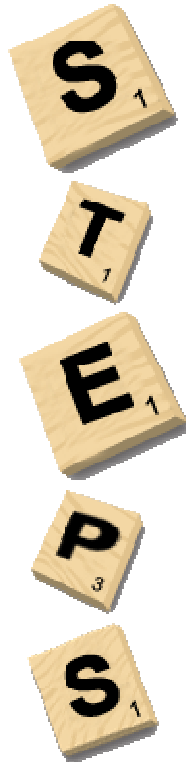
• flu burung	1922
• demam berdarah	1874
• kaki gajah	1648
• dinas kesehatan	1600
• rumah sakit	1417
• penyakit kaki	766
• berdarah dengue	656
• penderita kusta	658
• penyakit kusta	562
• kasus dbd	558

Term Frequency



Top 3-grams

Term Frequency



• penyakit kaki gajah	764
• demam berdarah dengue	639
• kepala dinas kesehatan	343
• penyakit demam berdarah	297
• dinas kesehatan kota	265
• virus flu burung	262
• dinas kesehatan kabupaten	251
• pemberantasan sarang nyamuk	241
• nyamuk aedes aegypti	240
• kasus flu burung	207



Hasil Pembangkitan Fitur

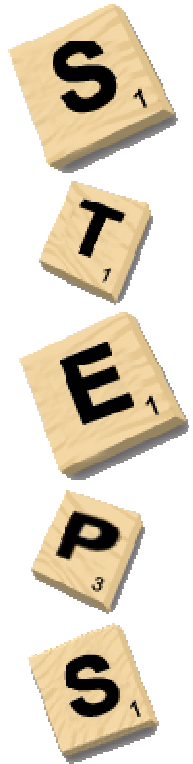


```
tropical_d 1:0.8333333333333333 2:0.1666666666666667
3:0 4:0 5:0 6:0 7:0.288461538461538
8:0.0769230769230769 9:0.0256410256410256 10:0
11:0.032051282051282 12:0.00641025641025641
13:0.178064516129032 14:0.0270967741935484
15:0.00387096774193548 16:0 17:0.0232258064516129
18:0 19:0.112087912087912 20:0.0263736263736264
21:0.00659340659340659 22:0 23:0.00659340659340659
24:0
```

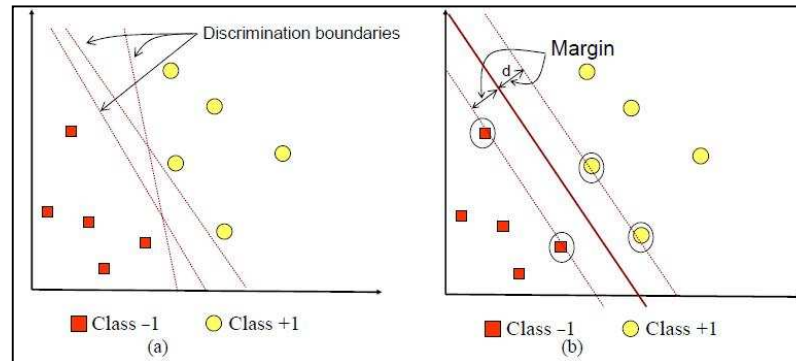
:
.
.
.
.

```
<title>satu warga tangerang suspect flu burung</title>
<content>ruang isolasi flu burung rsu tangerang kembali
merawat pasien suspect flu burung ...</content>
```

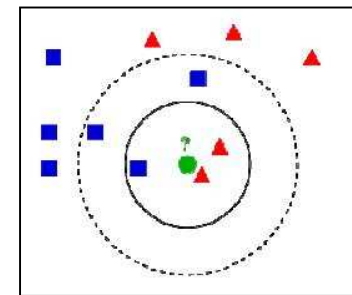
Metode Klasifikasi



- Support Vector Machine

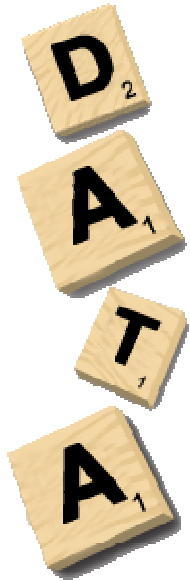


- KNN (k-Nearest Neighbors)



- Naïve Bayesian Classifier

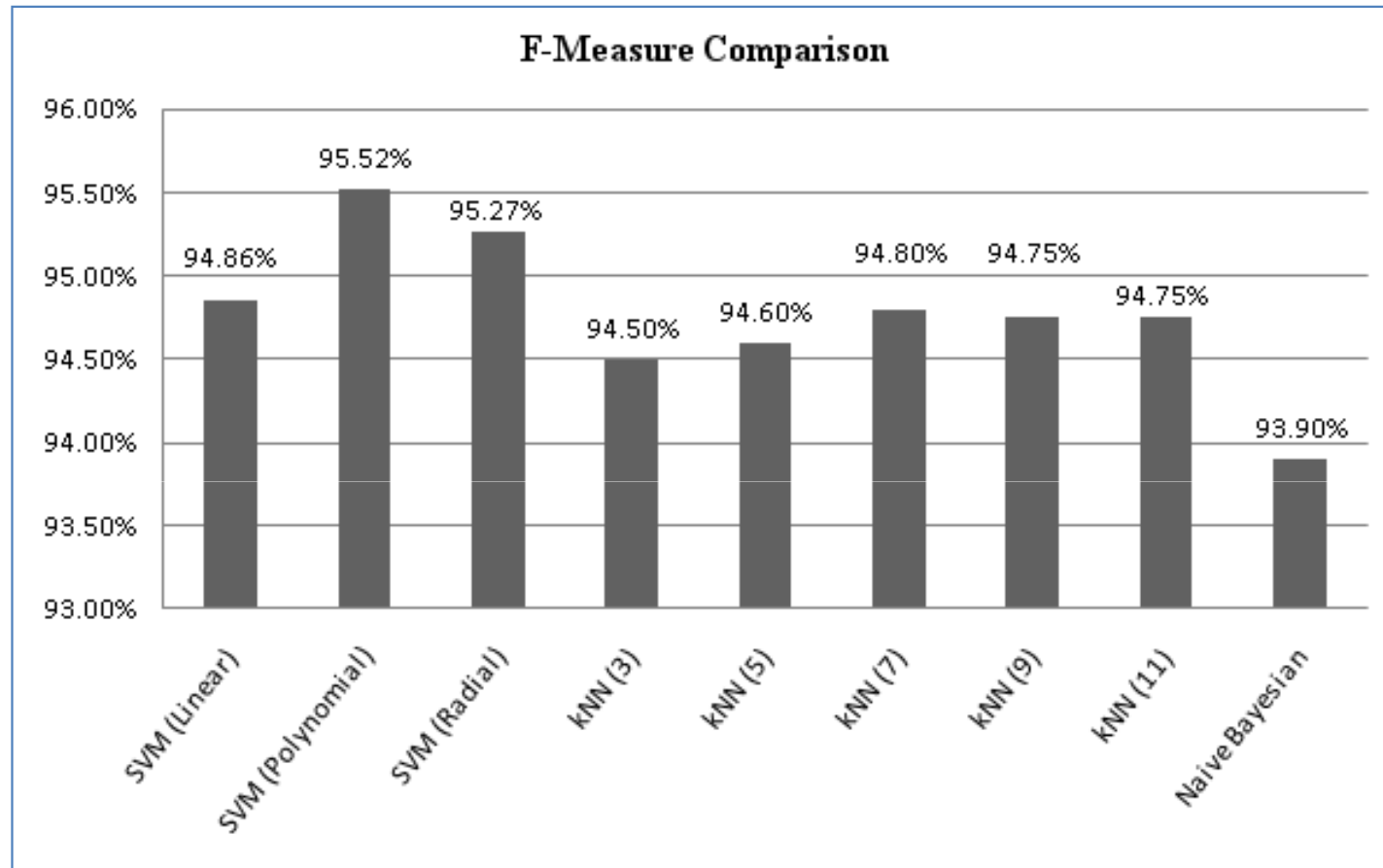
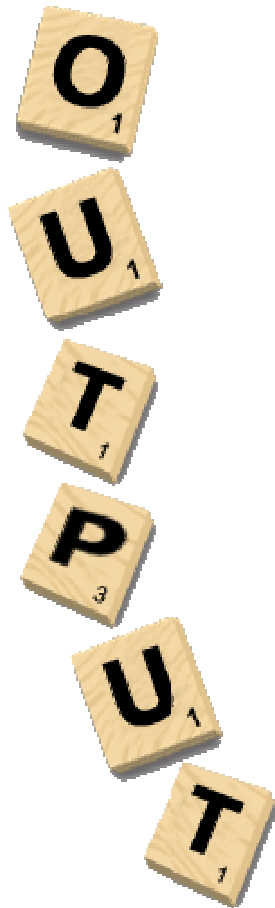
Data Percobaan



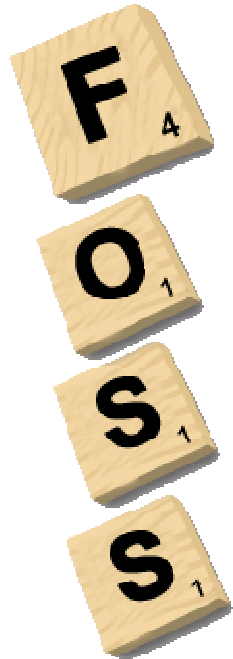
Category	Total		
	Documents	Training Set	Testing Set
Tropical Disease	1,857	1,300	557
Non-Tropical Disease	39,770	27,670	12,100
Total	41,627	28,970	12,657



Hasil



Open Source Tools



- Support Vector Machine

SVM^{light}

Support Vector Machine

Author: [Thorsten Joachims](mailto:thorsten@joachims.org) <thorsten@joachims.org>
[Cornell University](http://www.cornell.edu/~tj)
[Department of Computer Science](http://www.cornell.edu/~tj)

<http://svmlight.joachims.org/>

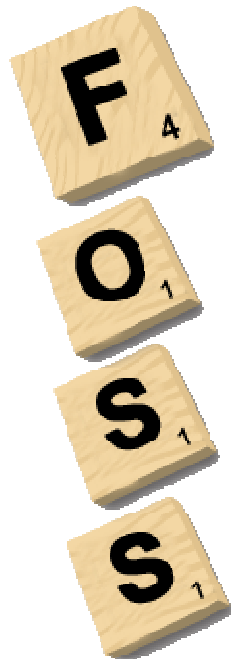
- KNN (k-Nearest Neighbors)
- Naïve Bayesian Classifier



<http://www.cs.waikato.ac.nz/~ml/weka/>



Open Source Tools



Waikato Environment for
Knowledge Analysis

Version 3.4.13

(c) 1999 - 2008
University of Waikato
New Zealand



GUI

Simple CLI

Explorer

Experimenter

KnowledgeFlow

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Undo | Edit... | Save...

Filter: Choose | None | Apply

Current relation:
Relation: iris
Instances: 150 Attributes: 5

Selected attribute:
Name: sepal.length Type: Numeric
Missing: 0 (0%) Distinct: 35 Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Attributes: All | None | Invert

No.	Name
1	<input type="checkbox"/> sepal.length
2	<input type="checkbox"/> sepal.width
3	<input type="checkbox"/> petal.length
4	<input type="checkbox"/> petal.width
5	<input type="checkbox"/> class

Remove

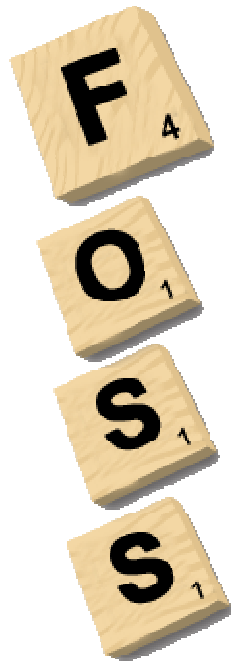
Class: class (Nom) | Visualize All

Status: OK | Log | x 0

<http://www.cs.waikato.ac.nz/~ml/weka/>



ARFF Data Format



```
File Edit View Search Terminal Help
@RELATION HB_Testing

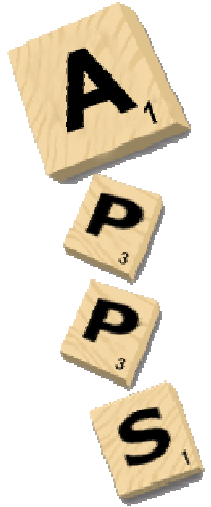
@ATTRIBUTE td_1    REAL
@ATTRIBUTE td_2    REAL
@ATTRIBUTE td_3    REAL
@ATTRIBUTE td_4    REAL
@ATTRIBUTE td_5    REAL
@ATTRIBUTE td_6    REAL
@ATTRIBUTE td_7    REAL
@ATTRIBUTE td_8    REAL
@ATTRIBUTE td_9    REAL
@ATTRIBUTE td_10   REAL
@ATTRIBUTE td_11   REAL
@ATTRIBUTE td_12   REAL
@ATTRIBUTE ntd_1   REAL
@ATTRIBUTE ntd_2   REAL
@ATTRIBUTE ntd_3   REAL
@ATTRIBUTE ntd_4   REAL
@ATTRIBUTE ntd_5   REAL
@ATTRIBUTE ntd_6   REAL
@ATTRIBUTE ntd_7   REAL
@ATTRIBUTE ntd_8   REAL
@ATTRIBUTE ntd_9   REAL
@ATTRIBUTE ntd_10  REAL
@ATTRIBUTE ntd_11  REAL
@ATTRIBUTE ntd_12  REAL
@ATTRIBUTE class   {tropical_d,non_tropical_d}

@DATA
0.75,0,0,0,0,0,0.264705882352941,0.0349264705882353,0.00551470588235294,0,0.0386029411764706,0.00367647058823529
,0.152918287937743,0.0221789883268483,0.00700389105058366,0,0.021011673151751,0.00233463035019455,0.113461538461
538,0.0375,0.00865384615384615,0,0.00769230769230769,0,tropical_d

1,1
```



Aplikasi 2: Akronim - Ekspansinya



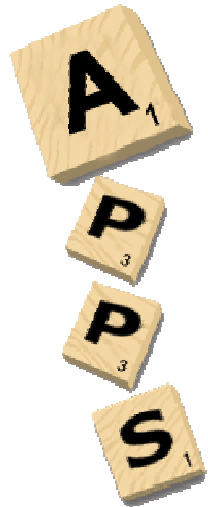
- *Penentuan Secara Otomatis Akronim dan Ekspansinya dari Data Teks Berbahasa Indonesia*
- Akronim pada umumnya didefinisikan sebagai singkatan yang dibentuk dari huruf awal atau gabungan huruf dari ekspansinya
- Database akronim dan kepanjangannya, sebagai contoh, dibutuhkan oleh mesin pencari untuk menentukan hasil pencarian yang relevan

Query:

Hasil **Ujian Nasional** 2011 \Leftrightarrow Hasil **UN** 2011



Aplikasi 2: Akronim - Ekspansinya



Seminar Nasional Teknologi Informasi & Komunikasi

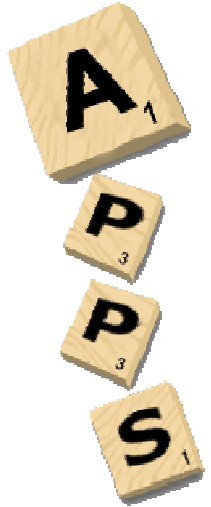
Pendahuluan



Dewasa ini, tidak dapat dipungkiri bahwa perkembangan teknologi sangat besar peranannya bagi kehidupan di jaman yang serba modern. Hal ini menjadikan teknologi sebagai kebutuhan vital, khususnya di kalangan dunia bisnis global. Dalam lingkup teknologi informatika, baik perangkat keras maupun perangkat lunak mengalami metamorfosa yang sangat cepat dan luar biasa. Terkait dengan perkembangan teknologi tersebut, khususnya Teknologi Open Source, Program Studi Teknik Informatika, Sekolah Tinggi Teknik Harapan (STTH) Medan mengadakan seminar yang bertaraf nasional. Kegiatan seminar ini diselenggarakan dalam rangka memperingati Hari Ulang Tahun Yayasan Pendidikan Harapan Medan yang ke-45. Kegiatan ini terselenggara atas kerjasama Program Studi Teknik Informatika STTH, Asosiasi Perguruan Tinggi Informatika dan Komputer (APTIKOM), Indonesian Mathematical Society (IndoMS), Aceh-Sumut dan Ikatan Pelajar dan Alumni Malaysia asal Sumatera Utara (IPAMSU) Sumut-Aceh, dengan tema "Pemanfaatan Teknologi Open Source sebagai Aplikasi dalam Bisnis Global".



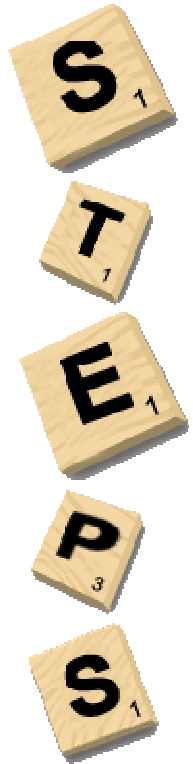
Riset Akronim Terkait



- J. Xu (2006): Menentukan akronim dan ekspansi menggunakan Support Vector Machine dan data adalah tulisan berbahasa Inggris. Hasil menunjukkan bahwa pendekatan machine learning lebih unggul dari pada pendekatan rule-based
- Chang (2002): Membangun kamus online dari akronim *biomedical* dan menentukan akronim dan ekspansi yang ditulis secara tidak teratur. Pasangan akronim dan kandidat ekspansi diberi skor berdasarkan kemiripan pasangan tersebut dengan data pembelajaran



Steps: Overview

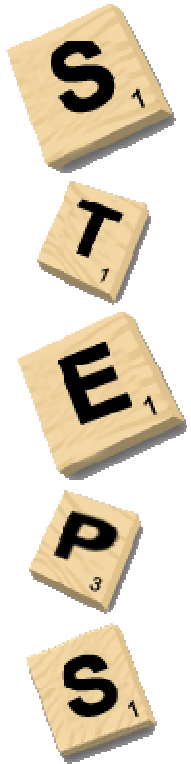


- Membangun kandidat akronim dan ekspansi
- Bila A diasumsikan sebagai sebuah kata dalam teks, maka A dianggap sebagai kandidat akronim jika jumlah huruf kapital dan angka (C_k) dalam kata A, dibagi karakter pembentuk kata A (C_s) melebihi threshold atau 75%

$$f(A) = \frac{\sum C_k}{\sum C_s} \geq 0.75$$

- Menbangun fitur dan skor numerik untuk setiap pasangan
- Metode KNN (k-Nearest Neighbors) sebagai metode klasifikasi

Steps: Fitur 1



- Fitur pertama (F_1) merepresentasikan hubungan antara jumlah karakter pada akronim dengan jumlah kata pada ekspansinya
- Jika A adalah karakter pembentuk akronim (huruf kapital ataupun non kapital) dan E adalah kata dalam ekspansi yang bukan kata penghubung dan kata depan maka F_1 :

$$F_1 = - \left[\left(\frac{\sum A}{\sum A + \sum E} {}^2 \log \frac{\sum A}{\sum A + \sum E} \right) + \left(\frac{\sum E}{\sum A + \sum E} {}^2 \log \frac{\sum E}{\sum A + \sum E} \right) \right]$$

- F_1 akan bernilai 1 jika jumlah karakter pembentuk akronim dan jumlah kata pembentuk ekspansi adalah sama dan akan bernilai < 1 jika sebaliknya

Steps: Fitur 2

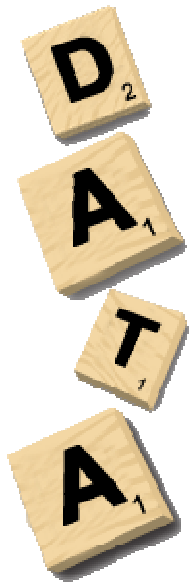


- Fitur kedua (F_2) merepresentasikan rasio kata pada ekspansi yang huruf awalnya ditulis dengan huruf kapital

$$F_2 = \frac{\sum E_K - \sum E_k}{|A|}$$

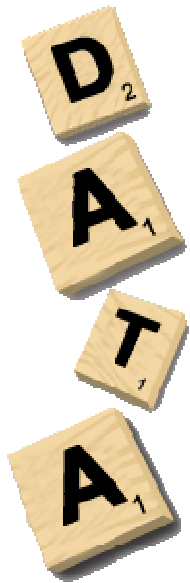
- Bank Syariah Mandiri** yang merupakan ekspansi dari **BSM** memiliki nilai $F_2 = 1$ karena jumlah huruf kapital pada awal kata ekspansi dan panjang akronim sama-sama bernilai 3 sementara jumlah huruf kecil pada awal kata ekspansi tidak ada. Namun untuk **Fakultas Matematika dan Ilmu Pengetahuan Alam** yang merupakan ekspansi dari **FMIPA**, fitur F_2 bernilai 0,80

Dataset: Data Pembelajaran



- Data pembelajaran (*training set*) sebanyak 4.570 pasangan akronim dan ekspansi yang dibangun dari 4 domain:
 - www.mediajurnalindonesia.com
 - www.okezone.com
 - www.republika.com
 - www.detik.com
- Dari 4.570 data pembelajaran, ditemukan sebanyak 54 pasangan dengan klas 1 (pasangan akronim dan ekspansi: **benar**), sedangkan 4.516 pasangan dengan klas 0 (pasangan akronim dan ekspansi: **salah**)

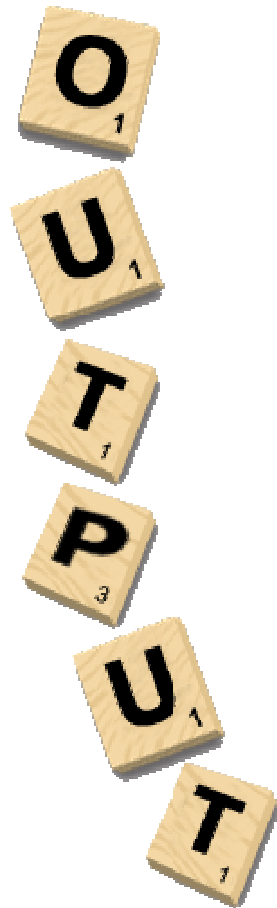
Dataset: Data Pengujian



- Data pengujian diambil secara acak dari artikel berita yang berkategori *teknologi, kesehatan, ekonomi, sosial, dan budaya*, sebanyak 50 web untuk setiap domain
- Data pengujian juga dibentuk dalam bentuk pasangan akronim & ekspansi, dari setiap domain diperoleh pasangan data sebagai berikut:

• Okezone.com	17.258
• Detik.com	15.731
• Media Jurnal Indonesia	18.970
• Republika.com	27.370

Hasil (*Results*)



Tabel 1. Hasil klasifikasi data Okezone.com

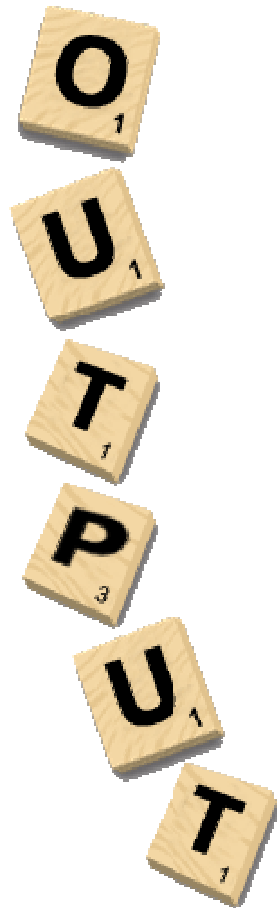
<i>k</i>	<i>Actual</i>	<i>Classified</i>		<i>Sum</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
		+	-				
<i>k</i> =3	+	84	29	17258	0,93	0,74	0,83
	-	6	17139				
<i>k</i> =5	+	84	29	17258	0,93	0,74	0,83
	-	6	17139				
<i>k</i> =7	+	84	29	17258	0,93	0,74	0,83
	-	6	17139				

Tabel 2. Hasil klasifikasi data Detik.com

<i>K</i>	<i>Actual</i>	<i>Classified</i>		<i>Sum</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
		+	-				
<i>k</i> =3	+	53	3	15731	0,96	0,95	0,95
	-	2	15673				
<i>k</i> =5	+	53	3	15731	0,96	0,95	0,95
	-	2	15673				
<i>k</i> =7	+	53	3	15731	0,96	0,95	0,95
	-	2	15673				



Hasil (*Results*)



Tabel 3. Hasil klasifikasi data Jurnal Indonesia

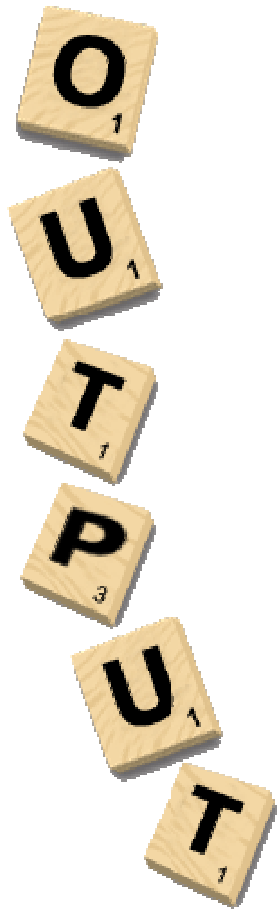
<i>k</i>	<i>Actual</i>	<i>Classified</i>		<i>Sum</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
		+	-				
<i>k</i> =3	+	81	9	18970	0,95	0,90	0,93
	-	4	18876				
<i>k</i> =5	+	81	9	18970	0,95	0,90	0,93
	-	4	18876				
<i>k</i> =7	+	81	9	18970	0,95	0,90	0,93
	-	4	18876				

Tabel 4. Hasil klasifikasi data Republika.com

<i>k</i>	<i>Actual</i>	<i>Classified</i>		<i>Sum</i>	<i>Precision</i>	<i>Recall</i>	<i>F-score</i>
		+	-				
<i>k</i> =3	+	103	23	27370	0,95	0,82	0,88
	-	5	27239				
<i>k</i> =5	+	103	23	27370	0,95	0,82	0,88
	-	5	27239				
<i>k</i> =7	+	103	23	27370	0,95	0,82	0,88
	-	5	27239				



Hasil (*Results*)

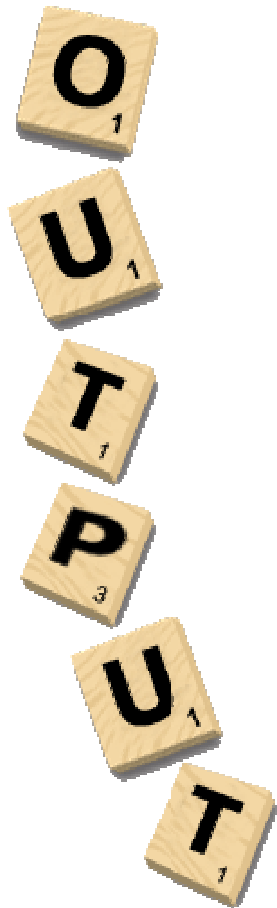


Tabel 5. Contoh akronim dan ekspansi yang benar

Akronim	Ekspansi
UN	Ujian Nasional
ATM	Anjungan Tunai Mandiri
BPM	Badan Pemberdayaan Masyarakat
FbPS	Facebook Prabowo Subianto
LP3I	Lembaga Pendidikan Pengembangan Profesi Indonesia
COAST	Cache On A Stick
CD-ROM	Compact Disk Read Only Memory
DPR	Dewan Perwakilan Rakyat
FeRAM	Ferroelectric Random Access Memory



Hasil (*Results*)



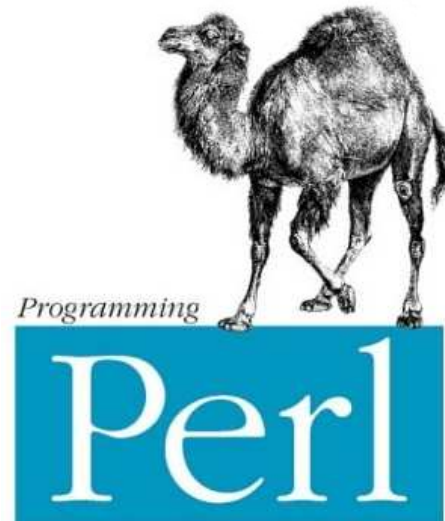
Tabel 6. Contoh akronim dan ekspansi yang salah

Akronim	Ekspansi
HMSP	HM Sampoerna
PTBA	PT Tambang Batubara Bukit Asam
BPMIGAS	Badan Pelaksana Kegiatan Usaha Hulu Minyak dan Gas Bumi
PT	perekonomian tambahnya
TPI	tangkapan ikan
KSSK	khususnya mengenai keuangan katanya
ITB	internasional yang bersifat
PKP2B	pertambangan KP BUMN

Algoritma gagal menebak akronim dan ekspansi yang terbentuk dari gabungan karakter kata tetapi gabungan kata tersebut tidak ditulis menggunakan huruf kapital seperti Panja, Bappeda, atau Unsyiah. Namun, algoritma dapat menentukan akronim-akronim tersebut bila akronim tersebut ditulis menjadi PANJA, BAPPEDA, atau UNSYIAH



www.informatika.unsyiah.ac.id/tfa



CGI programming