CS 105

Week 7

MP4: Canceled

Weekly Reading #6

MP1 – MP3 Regrades

Data

All data is made up of 0s and 1s.

File extensions tell us how to interpret the 0s and 1s.

File Extensions

- The part of the file **after the dot**:

File Extensions

- The part of the file **after the dot**:

  .js: JavaScript code

File Extensions

- The part of the file **after the dot**:

  .js: JavaScript code
  .html: HyperText Markup Language

File Extensions

- The part of the file **after the dot**:

  .js: JavaScript code
  .html: HyperText Markup Language
  .xlsx: Excel

File Extensions

- The part of the file **after the dot**:

  .js: JavaScript code
  .html: HyperText Markup Language
  .xlsx: Excel
  .xml: eXtensible Markup Language

File Extensions

- The part of the file **after the dot**:

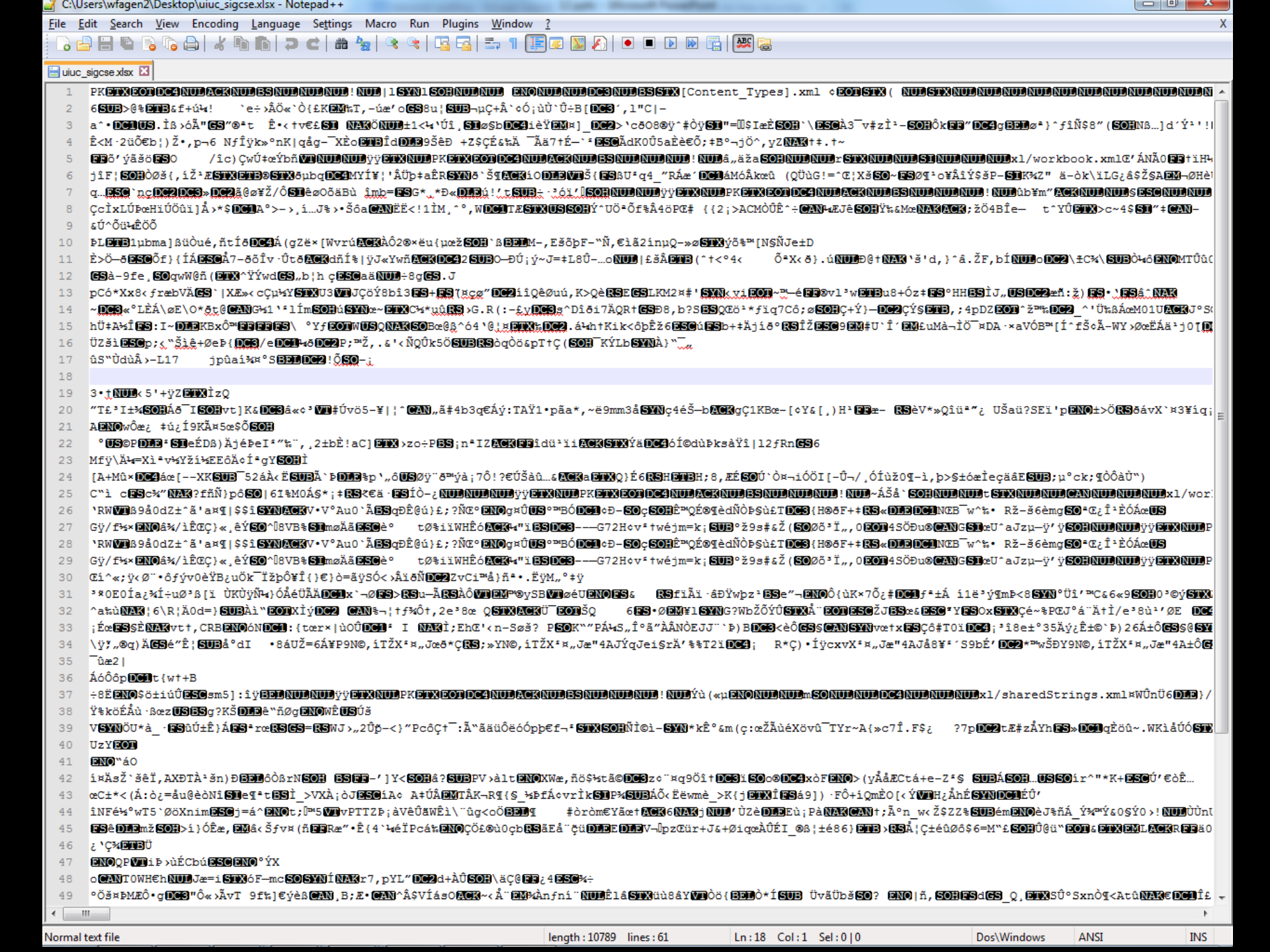
  .js: JavaScript code
  .html: HyperText Markup Language
  .xlsx: Excel
  .xml: eXtensible Markup Language
  .csv: Comma Separated Values

File Extensions

- The part of the file **after the dot**:

  .js: JavaScript code
  .html: HyperText Markup Language
  .xlsx: Excel
  .xml: eXtensible Markup Language
  .csv: Comma Separated Values
  .fasta: DNA Data

Some file types do not make sense outside of their intended context.


C:\Users\wfagen2\Desktop\uiuc_sigcse.xlsx - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
uiuc_sigcse.xlsx
Normal text file
length : 10789 lines : 61
Ln : 18 Col : 1 Sel : 0 | 0
Dos\Windows
ANSI
INS

API

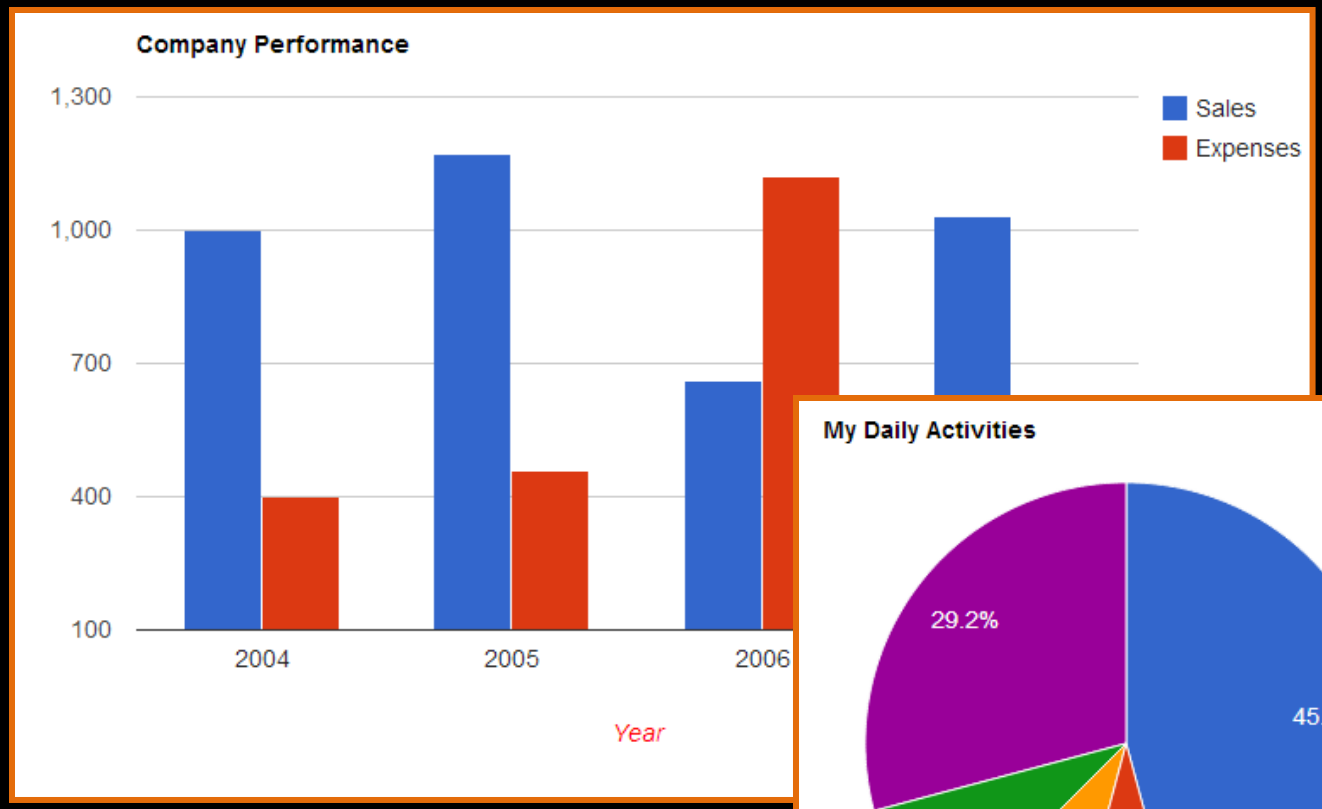
API

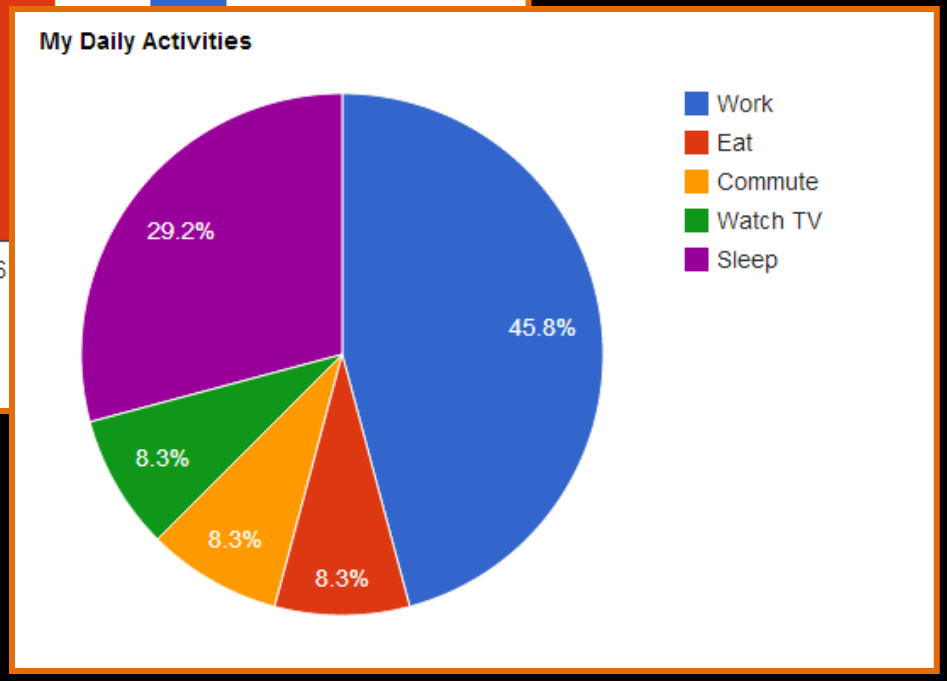
Application Programming Interfaces

APIs

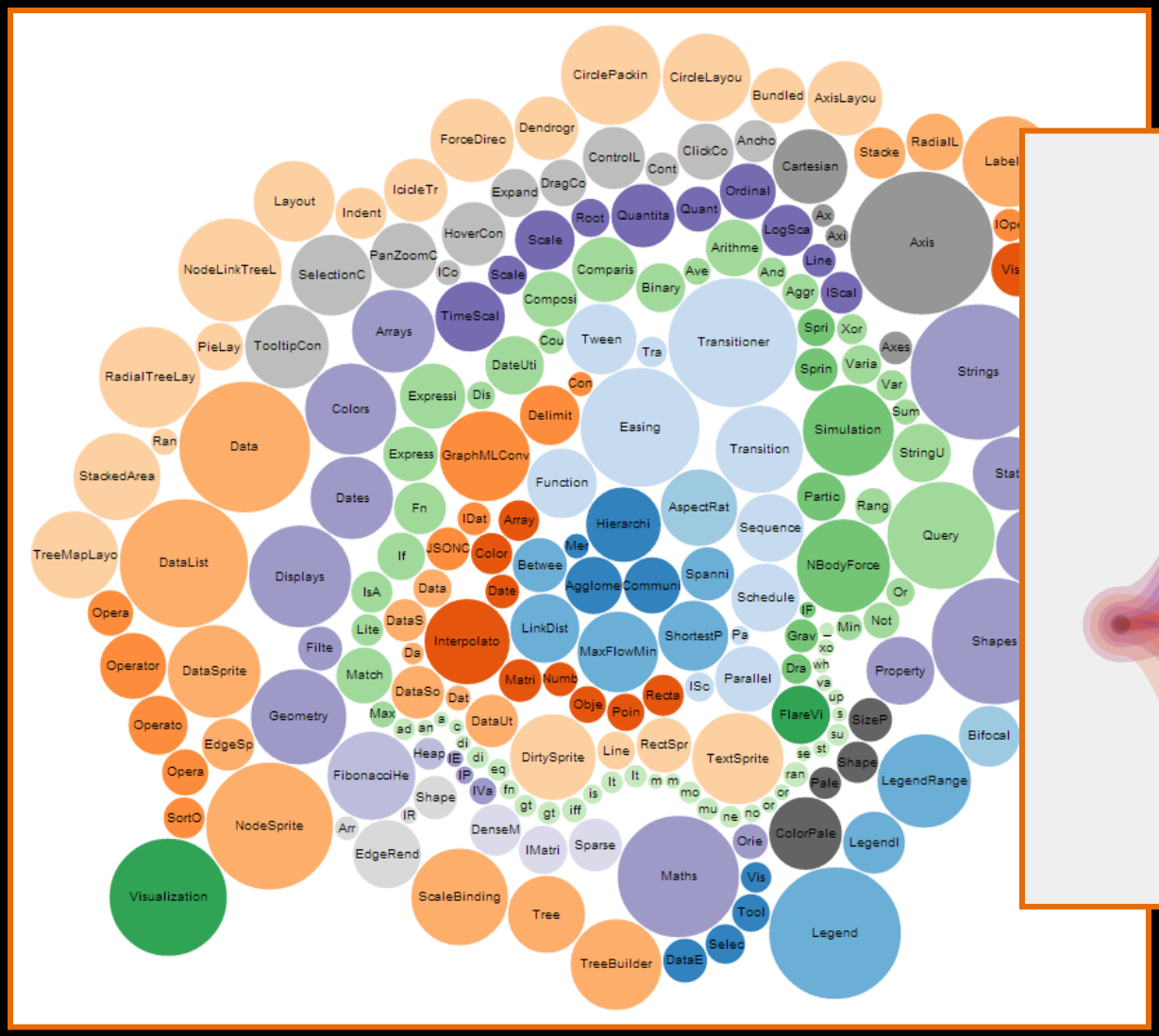
- Allows a programmer (like you) to access functionality or data from someone else.

Google Graphs API





d3


CirclePackin
CircleLayou
Bundled
AxisLayou
ForceDirec
Dendrogr
ControlL
ClickCo
Ancho
Cartesian
Stacke
RadialL
Label
Layout
Indent
IcicleTr
Expand
DragCo
Cont
Ordinal
HoverCon
Root
Quantita
Quant
Axis
NodeLinkTreeL
SelectionC
PanZoomC
Scale
Arithme
LogSca
Ax
Axi
Line
Comparis
Binary
Ave
And
Aggr
IScal
TimeScal
Composi
Arrays
PieLay
TooltipCon
Tween
Transitioner
Spri
Xor
Axes
RadialTreeLay
DateUti
Sprin
Varia
Var
Strings
Colors
Expressi
Dis
Delimit
Easing
Sum
Ran
Data
Express
GraphMLConv
Simulation
StringU
StackedArea
Dates
Function
Transition
Fn
IDat
Array
AspectRat
Partic
Rang
Query
TreeMapLayo
DataList
If
JSONC
Color
Betwee
Hierarchi
Sequence
Displays
IsA
Data
Agglome
Communi
Spanni
NBodyForce
Opera
Date
Schedule
Or
Operator
DataSprite
Lite
DataS
Interpolato
LinkDist
MaxFlowMin
ShortestP
Pa
Min
Not
Shapes
Filte
Da
Grav
Match
Matri
Numb
Parallel
Property
Operato
Max
DataSo
Dat
Obje
Poin
Recta
ISc
Dra
FlareVi
SizeP
Geometry
DataUt
EdgeSp
Heap
DirtySprite
Line
RectSpr
TextSprite
Shape
Bifocal
Opera
FibonacciHe
Pale
LegendRange
SortO
NodeSprite
Shape
DenseM
ColorPale
Legendl
Arr
EdgeRend
IMatri
Sparse
Maths
Orie
Visualization
ScaleBinding
Tree
Vis
Tool
Legend
TreeBuilder
DataE
Selec
General science and ba...
Space flight (253)
Space, science, applic...
Supporting space activ...
Research and general e...
Training and employmen...
Preprimary thru second...
Other labor services (...
Pensions
Tertiary education
Social services (506)
Health Care
Education not definabl...
Education
Current Operations Ed...
Subsidiary services to...
Defense
Current Operations St...
R&D Education
Welfare
Current Operations Li...
Education n.e.c.
Protection
Construction State Ed...
Transportation
Construction Librarie...
General Government
Other Capital Outlay ...
Other Spending
Other Capital Outlay ...
Interest
Assistance and Subsidi...
Balance
Total Spending
Federal Deficit
Gross Public Debt

Data Formats

XML

XML

eXtensible Markup Language

XML

- Very similar to HTML:

XML

- Very similar to HTML:
  - Consists of three types of tags:
    start tags: <section>

XML

- Very similar to HTML:
  - Consists of three types of tags:
    start tags: <section>
    end tags: </section>

XML

- Very similar to HTML:
  - Consists of three types of tags:
    - start tags: <section>
    - end tags: </section>
    - empty-element tags: <section />

XML

- Very similar to HTML:
  - Consists of three types of tags:
    start tags: <section>
    end tags: </section>
    empty-element tags: <section />
  - All start tags must have an end tag.

XML Example

- Yahoo! Currency API:
  http://finance.yahoo.com/webservice/v1/symbols/allcurrencies/quote

CSV

CSV

Comma Separated Values

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV

- Related values are listed on one line, separated by commas:

12,14,16,18,20,22,24

CSV Example

- Yahoo! Finance API:
  http://finance.yahoo.com/d/quotes.csv?s=AAPL+GOOG+MSFT&f=nab

CSV Example

- Yahoo! Finance API:
  http://finance.yahoo.com/d/quotes.csv?s=AAPL+GOOG+MSFT&f=nab

```
"Apple Inc.",529.31,528.61
"Google Inc.",1214.90,1214.21
"Microsoft Corpora",37.83,37.78
```

CSV

- Excel will read CSV files directly!

CSV

- Excel will read CSV files directly!

  …so we will often try and get our data in CSV, if it is available.

Application Specific Data

Application Specific Data

- Many applications use their own data formats.

Application Specific Data

- Many applications use their own data formats.

  DNA Data uses the **FASTA file format**

Application Specific Data

- Many applications use their own data formats.

  DNA Data uses the **FASTA file format**

```
>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP
```

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A

\- gap of indeterminate length

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A

\- gap of indeterminate length

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

>gi|129295|sp|P01013|OVAX_CHICK GENE X PROTEIN (OVALBUMIN-RELATED)
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNV
ATLPAEKMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLP
QMKIEEKYNLTSVLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGV
IEDIKHSPESEQFRADHPFLFLIKHNPTNTIVYFGRYWSP

VSSST

A	adenosine	C	cytidine
G	guanine	T	thymidine
N	A/G/C/T (any)	U	uridine
K	G/T (keto)	S	G/C (strong)
Y	T/C (pyrimidine)	M	A/C (amino)
W	A/T (weak)	R	G/A (purine)
B	G/T/C	D	G/A/T
H	A/C/T	V	G/C/A
-	gap of indeterminate length		

Excel