

Python in NLP

Nattapol KRITSUTHIKUL
NECTEC, Thailand

<https://1st.nectec.or.th/nattapol/private/RMUTT-2022-AI/>

Developer Tools

- Python 3.8 or above
 - <https://www.python.org/downloads/>
- Offline
 - JetBrains PyCharm
 - Jupyter Notebooks
 - Visual Studio Code
 - Spyder
- Online
 - Google Colaboratory (Colab) [GPU]
 - Microsoft Azure Notebooks [GPU]
 - Kaggle Notebooks / Kaggle Kernels [GPU]
 - Binder
 - CoCalc
 - JetBrains Datalore
 - Deepnote
 - ...

Create Project - Python 3.3 Virtual Environment

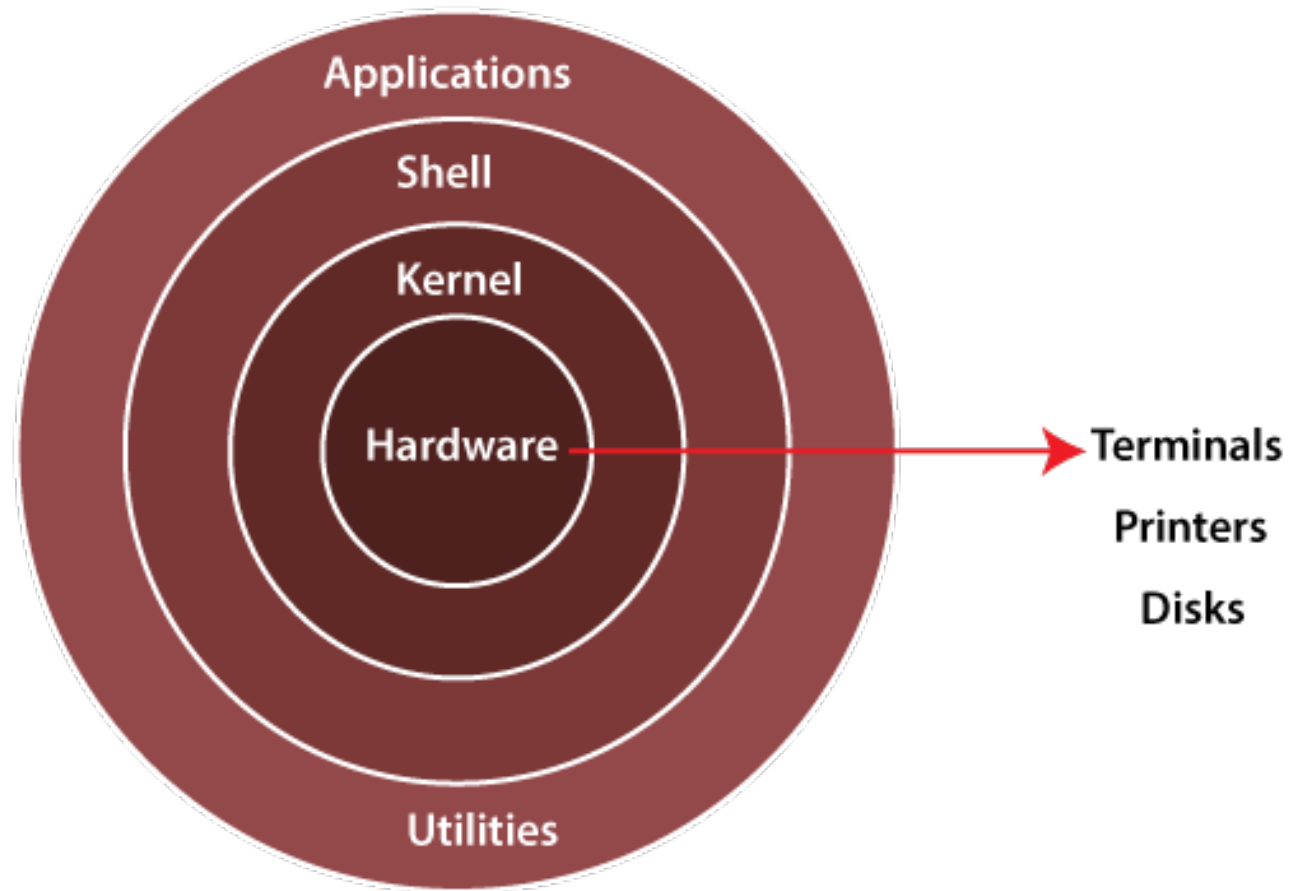
- Windows
 - `python -m venv myproject`
 - `myproject/Scripts/activate`
 - `myproject/Scripts/deactivate`
- Windows with Python Launcher
 - `py -m venv myproject`
 - `myproject/Scripts/activate`
 - `myproject/Scripts/deactivate`
- macOS / linux
 - `python3 -m venv ~/myproject`
 - `source ~/myproject/Scripts/activate`
 - `~/myproject/Scripts/deactivate`

Create Project - Anaconda

- Windows
 - `conda create -n myproject python=3.8`
 - `conda activate myproject`
 - `conda deactivate myproject`
- macOS / linux
 - `conda create -n ~/myproject python=3.8`
 - `conda activate myproject`
 - `conda deactivate myproject`

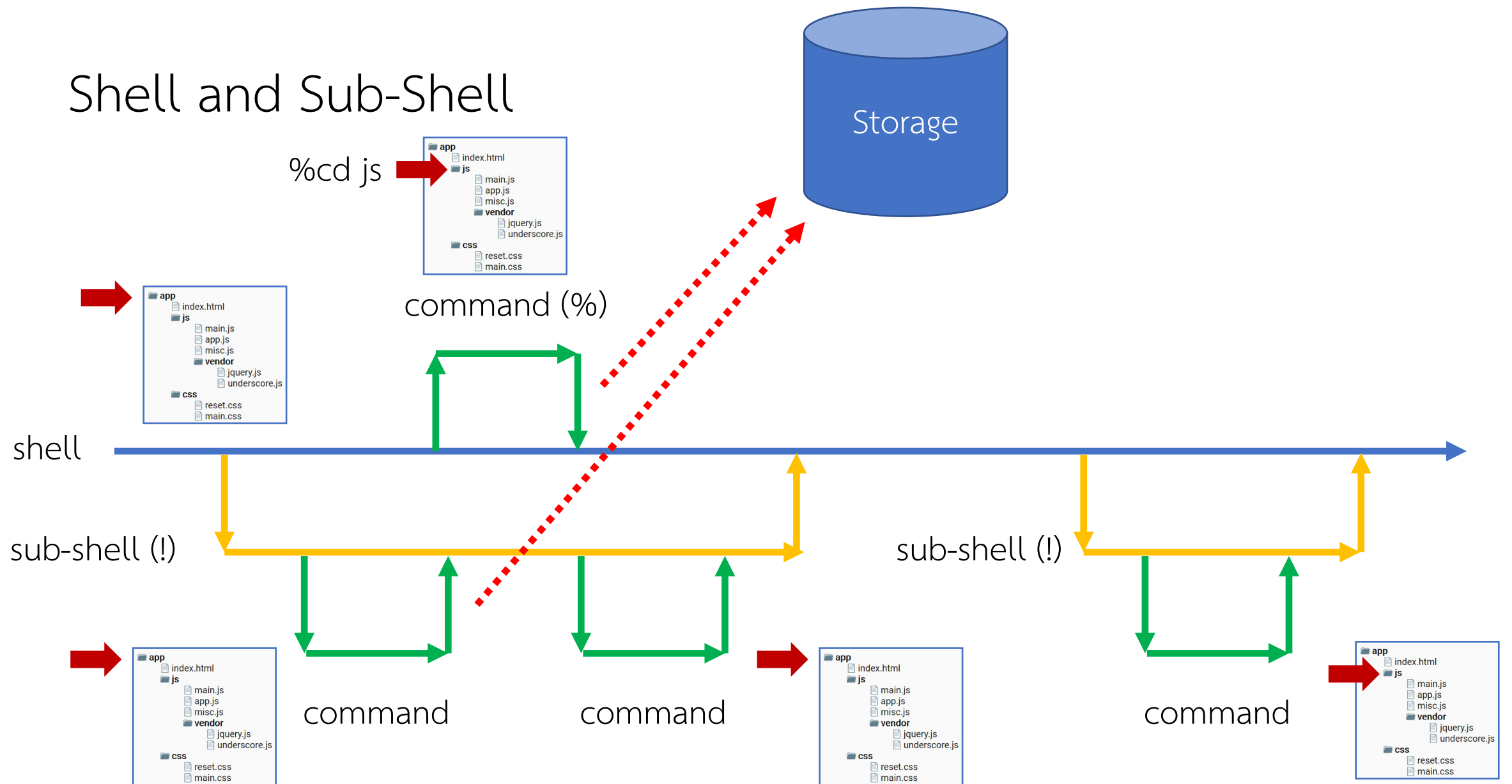
Workshop

Architecture of Modern Operating System



<https://www.javatpoint.com/architecture-of-linux>

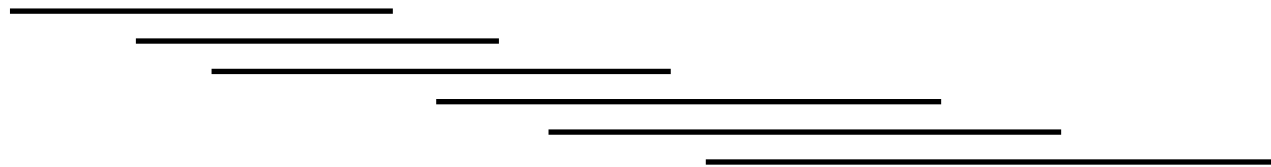
Shell and Sub-Shell



n -gram

n -gram = 3

S = she be always the most fabulous pop singer



n -gram

- What n -gram can tell us?
 - How possible a sentence can be segmented into consecutive phrases
 - How possible a sentence can be formed in the target language

n -gram

- Approximation of word sequence probability

$$\begin{aligned} P(w_1 w_2 w_3 \dots w_n) &= P(w_1) \times P(w_2 \mid w_1) \times P(w_3 \mid w_1 w_2) \times \dots \times P(w_n \mid w_1 w_2 \dots w_{n-1}) \\ &= \prod_{k=1}^n P(w_k \mid w_1 w_2 \dots w_{k-1}) \\ &\approx \prod_{k=1}^n P(w_k) && \text{unigram (1-gram)} \\ &\approx \prod_{k=1}^n P(w_k \mid w_{k-1}) && \text{bigram (2-gram)} \\ &\approx \prod_{k=1}^n P(w_k \mid w_{k-2} w_{k-1}) && \text{trigram (3-gram)} \\ &\approx \prod_{k=1}^n P(w_k \mid w_{k-x} \dots w_{k-2} w_{k-1}) && x\text{-gram (generalized)} \end{aligned}$$

Challenge #1

Student	Exam1	Exam2	Exam3
John	9	9	7
Maria	7	8	9
Julia	2	5	7

Exam1

1: John
2: Maria
3: Julia
AVG: 6
Min: 2

Exam2

1: John
2: Maria
3: Julia
AVG: 7.33
Min: 5

Exam3

1: Maria
2: John, Julia
AVG: 7.66
Min: 7

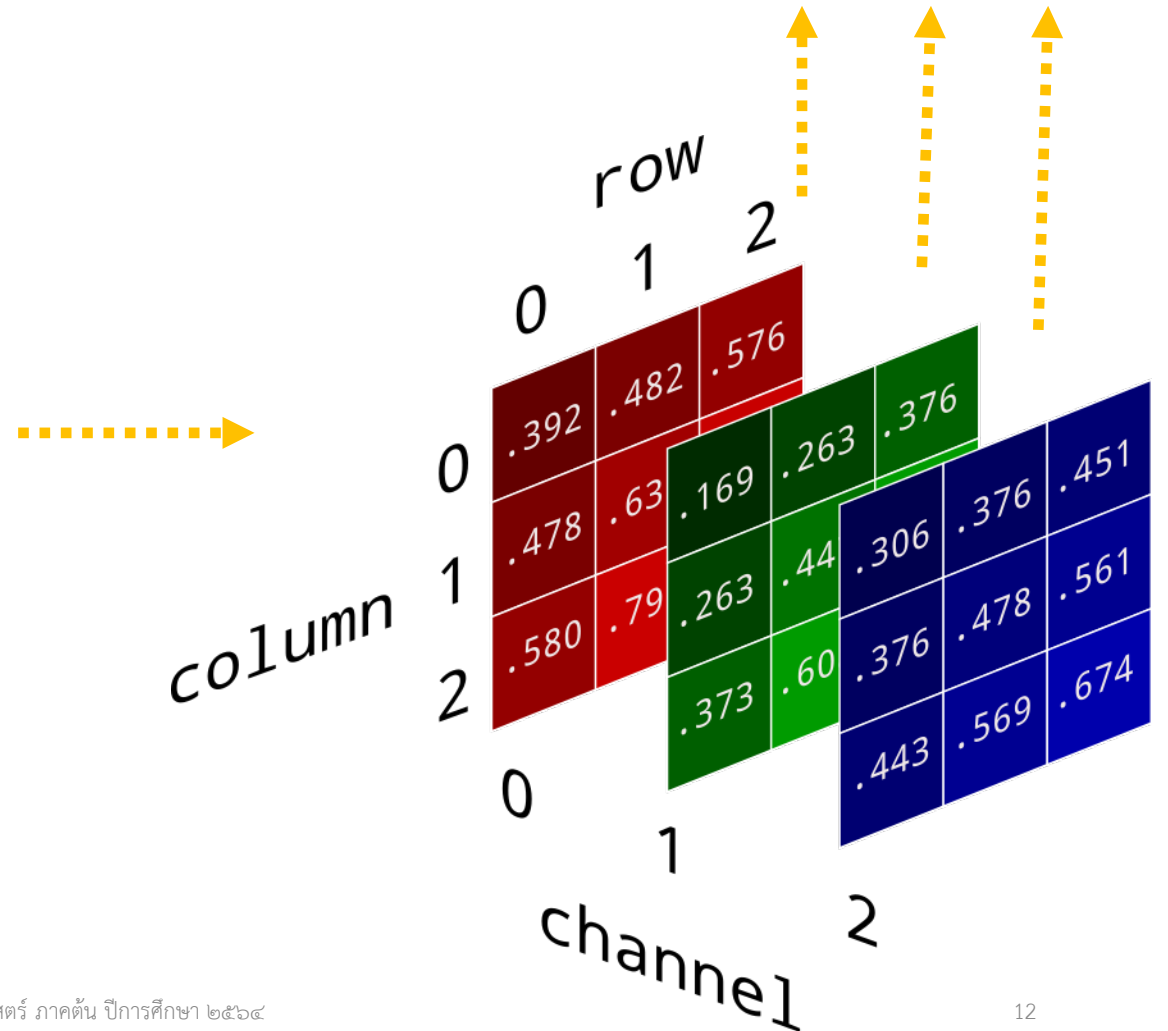
Total Score

1: John (25 points)
2: Maria (24 points)
3: Julia (14 points)

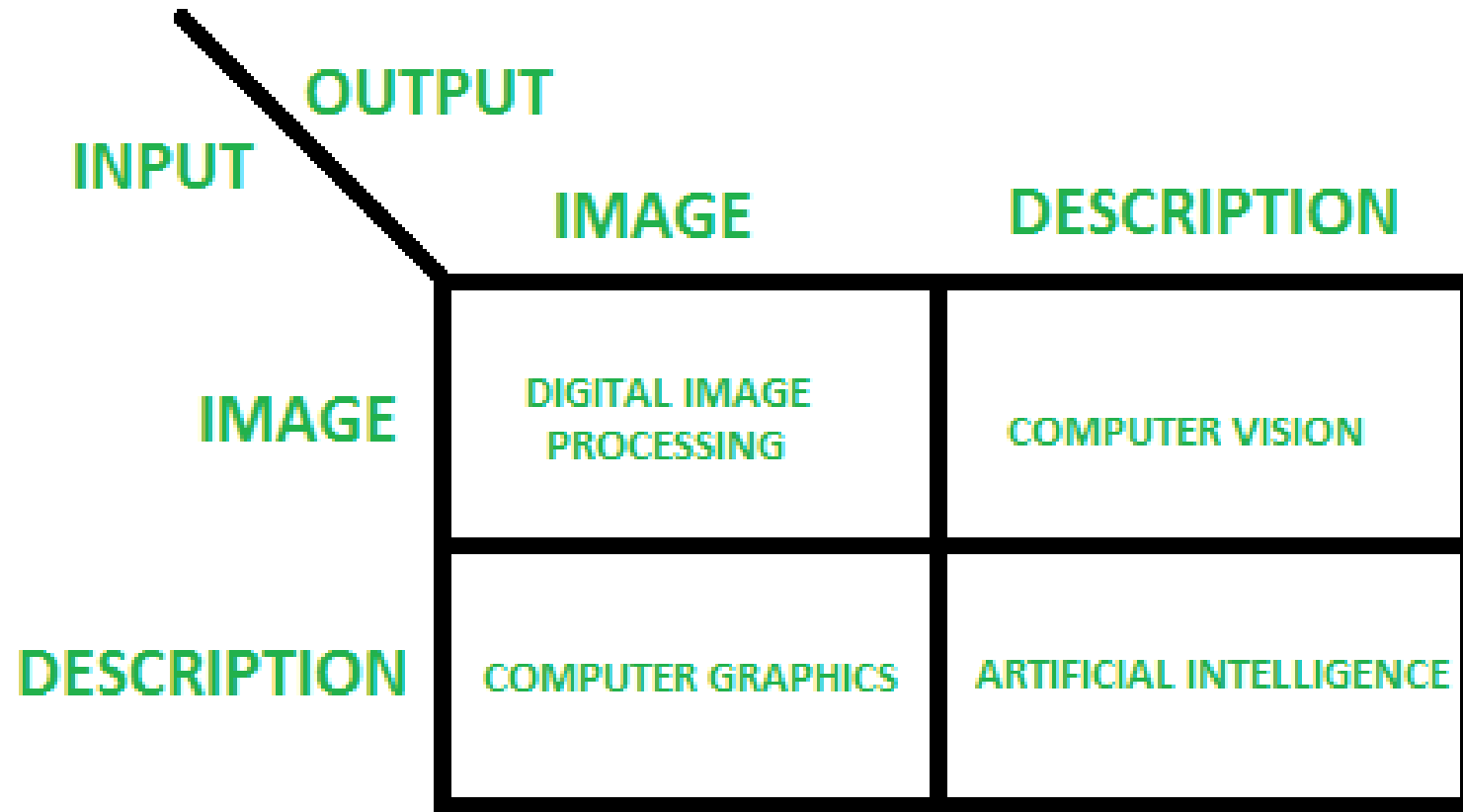
(Digital) Image Processing

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & f(0,2) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & f(1,2) & \dots & f(1,N-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & f(M-1,2) & \dots & f(M-1,N-1) \end{bmatrix}$$

$$\leftarrow \dots f(x,y) = \begin{bmatrix} r(x,y) \\ g(x,y) \\ b(x,y) \end{bmatrix}$$



Overlapping Fields in Image Processing

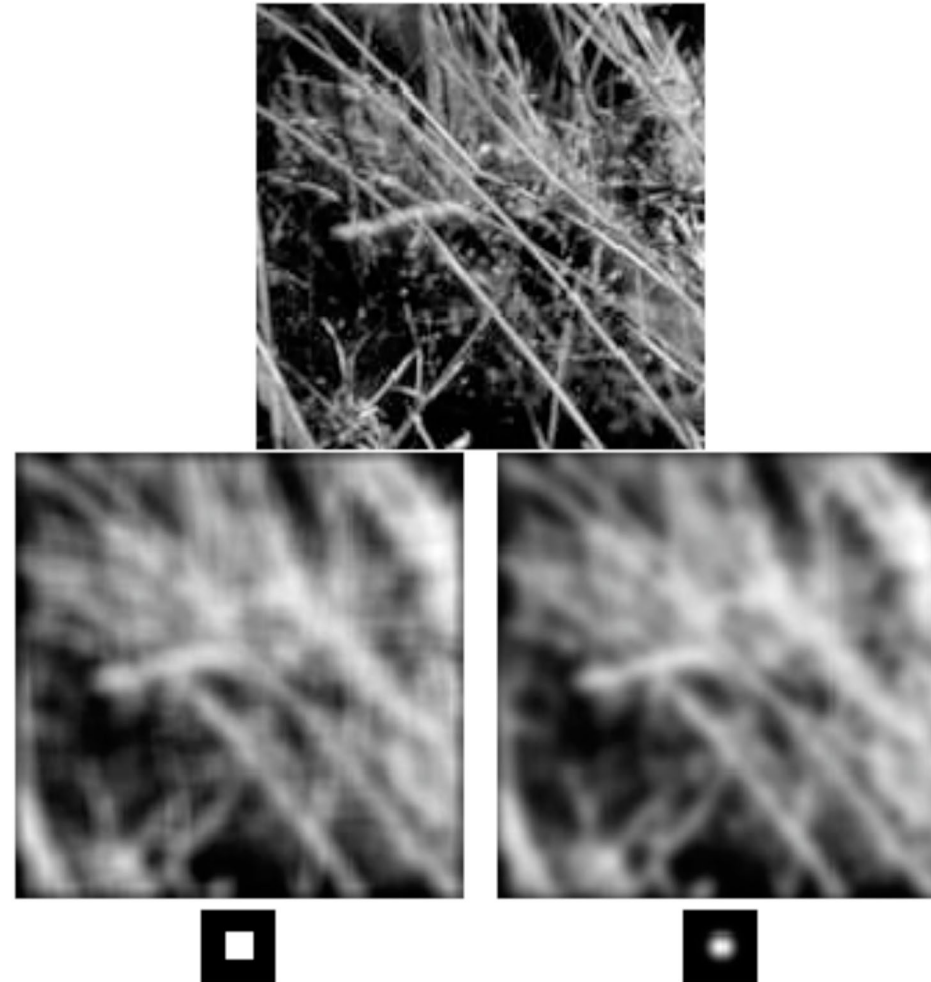


Source: Digital Image Processing (Rafael c. gonzalez)

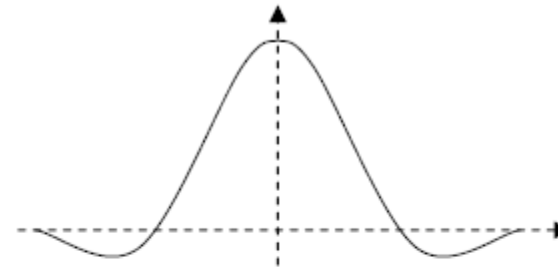
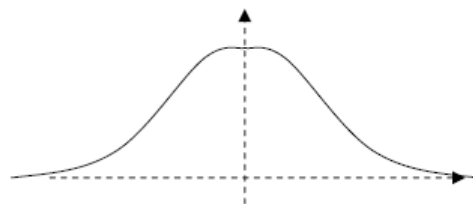
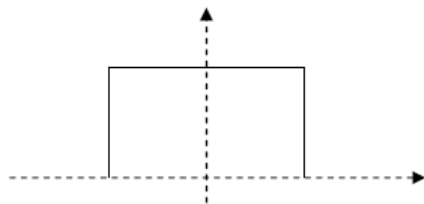
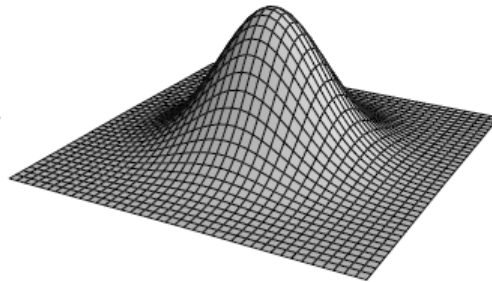
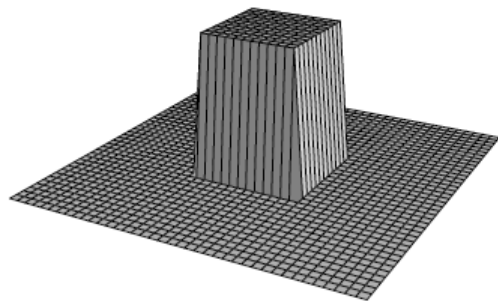
Blur Filter



Box Filter vs Gaussian Filter



Box Filter, Gaussian Filter, and Laplace Filter




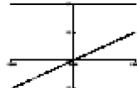
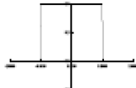

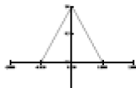

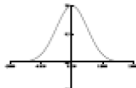


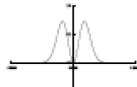
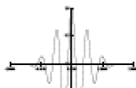
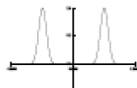
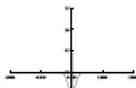
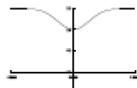




0	0	0	0	0
0	1	1	1	0
0	1	1	1	0
0	1	1	1	0
0	0	0	0	0

0	1	2	1	0
1	3	5	3	1
2	5	9	5	2
1	3	5	3	1
0	1	2	1	0

0	0	-1	0	0
0	-1	-2	-1	0
-1	-2	16	-2	-1
0	-1	-2	-1	0
0	0	-1	0	0

Some Examples

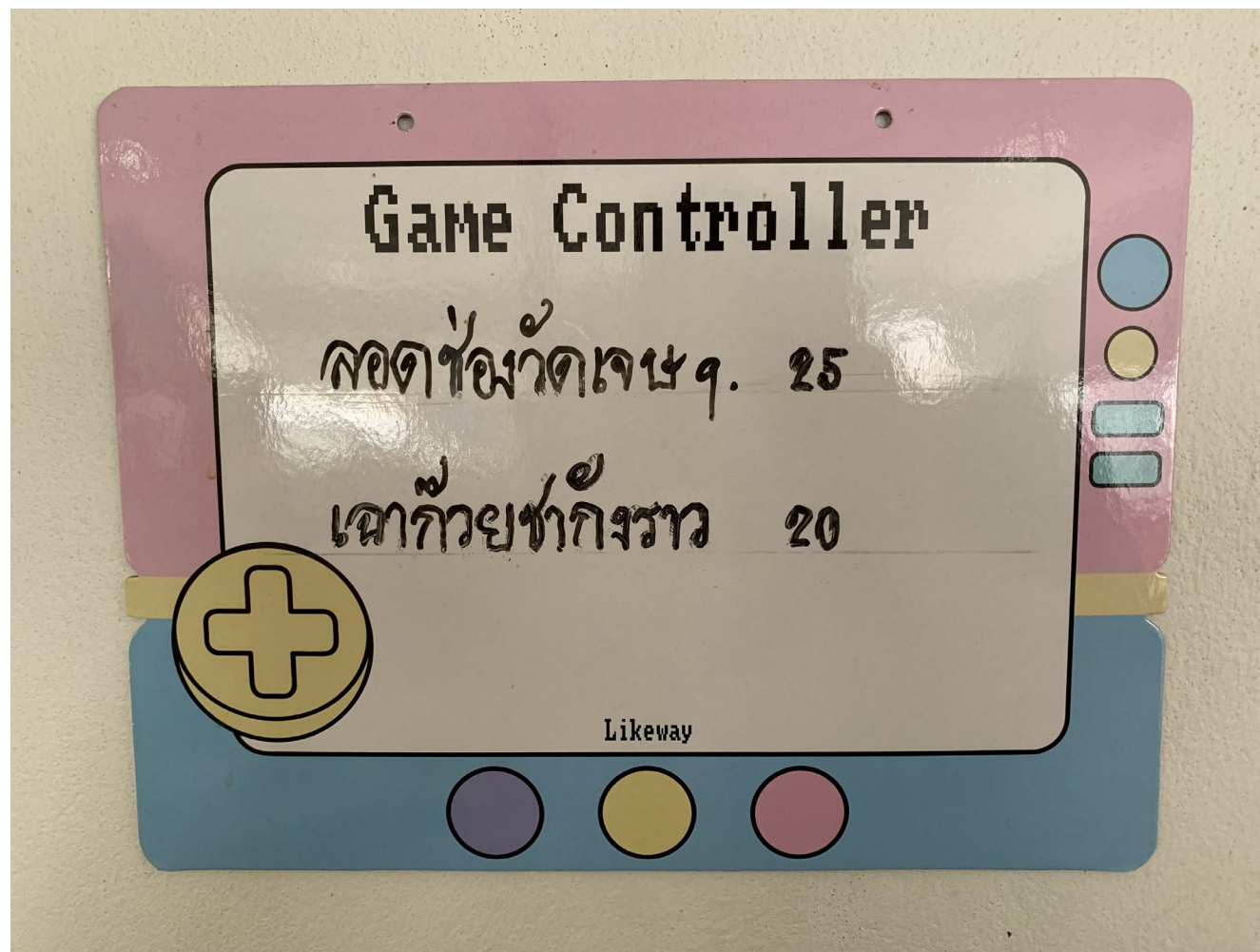
Name	Signal	Signal	Transform	Transform
impulse		$\delta(x)$	\Leftrightarrow	1 
shifted impulse		$\delta(x - u)$	\Leftrightarrow	$e^{-j\omega u}$ 
box filter		$\text{box}(x/a)$	\Leftrightarrow	$a\text{sinc}(a\omega)$ 
tent		$\text{tent}(x/a)$	\Leftrightarrow	$a\text{sinc}^2(a\omega)$ 
Gaussian		$G(x; \sigma)$	\Leftrightarrow	$\frac{\sqrt{2\pi}}{\sigma} G(\omega; \sigma^{-1})$ 
Laplacian of Gaussian		$(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})G(x; \sigma)$	\Leftrightarrow	$-\frac{\sqrt{2\pi}}{\sigma} \omega^2 G(\omega; \sigma^{-1})$ 
Gabor		$\cos(\omega_0 x)G(x; \sigma)$	\Leftrightarrow	$\frac{\sqrt{2\pi}}{\sigma} G(\omega \pm \omega_0; \sigma^{-1})$ 
unsharp mask		$(1 + \gamma)\delta(x) - \gamma G(x; \sigma)$	\Leftrightarrow	$(1 + \gamma) - \frac{\sqrt{2\pi}\gamma}{\sigma} G(\omega; \sigma^{-1})$ 
windowed sinc		$\text{rcos}(x/(aW)) \text{sinc}(x/a)$	\Leftrightarrow	(see Figure 3.29) 

Challenge #2 – How many stars in the picture?

<https://www.universetoday.com/34034/messier-35/>



What are the main contents in this picture?



License

Computer Software Overview

*** เนื้อหารวบรวมจาก DIP, TIS, SIPA และ NECTEC

- Source Code
 - ต้นฉบับโปรแกรมคอมพิวเตอร์
 - มนุษย์อ่านเข้าใจได้
 - คอมพิวเตอร์ไม่เข้าใจ
- Binary / Executable (.exe, .msi)
 - ภาษาที่คอมพิวเตอร์ใช้เพื่อทำงาน
 - มนุษย์อ่านไม่เข้าใจ
 - คอมพิวเตอร์อ่านเข้าใจ

```
00000000 fc 31 c0 8e c0 8e d8 8e d0 bc 00 7c 89 e6 bf 00 |.l.....|...|
00000010 06 b9 00 01 f3 a5 89 fd b1 08 f3 ab fe 45 f2 e9 |.....E..|
00000020 00 8a f6 46 bb 20 75 08 84 d2 78 07 80 4e bb 40 |...F. u...x..N.@|
00000030 8a 56 ba 88 56 00 e8 fc 00 52 bb c2 07 31 d2 88 |.V..V....R...l..|
00000040 6f fc 0f a3 56 bb 73 19 8a 07 bf 87 07 b1 03 f2 |o...V.s.....|
00000050 ae 74 0e b1 0b f2 ae 83 c7 09 8a 0d 01 cf e8 c5 |.t.....|
00000060 00 42 80 c3 10 73 d8 58 2c 7f 3a 06 75 04 72 05 |.B...s.X...u.r..|
00000070 48 74 0d 30 c0 04 b0 88 46 b8 bf b2 07 e8 a6 00 |Ht.0....F.....|
00000080 be 7b 07 e8 b2 00 8a 56 b9 4e e8 8e 00 eb 05 b0 |.{....V.N.....|
00000090 07 e8 b0 00 30 e4 cd 1a 89 d7 03 7e bc b4 01 cd |....0.....~...|
000000a0 16 75 0d 30 e4 cd 1a 39 fa 72 f2 8a 46 b9 eb 16 |.u.0...9.r..F...|
000000b0 30 e4 cd 16 88 e0 3c 1c 74 f1 2c 3b 3c 04 76 06 |0.....<.t...<.v.|
000000c0 2c c7 3c 04 77 c9 98 0f a3 46 0c 73 c2 88 46 b9 |...<.w...F.s..F.|
000000d0 be 00 08 8a 14 89 f3 3c 04 9c 74 0a c0 e0 04 05 |.....<.t.....|
000000e0 be 07 93 c6 07 80 53 f6 46 bb 40 75 08 bb 00 06 |.....S.F.@u...|
000000f0 b4 03 e8 59 00 5e 9d 75 06 8a 56 b8 80 ea 30 bb |...Y.^..u..V...0.|
00000100 00 7c b4 02 e8 47 00 72 86 81 bf fe 01 55 aa 0f |.|...G.r....U..|
00000110 85 7c ff be 85 07 e8 19 00 ff e3 b0 46 e8 24 00 |.|.....F.$..|
00000120 b0 31 00 d0 eb 17 0f ab 56 0c be 78 07 e8 eb ff |.l.....V..x...|
00000130 89 fe e8 03 00 be 85 07 ac a8 80 75 05 e8 04 00 |.....u...|
00000140 eb f6 24 7f 53 bb 07 00 b4 0e cd 10 5b c3 8a 74 |..$.S.....[.t|
00000150 01 8b 4c 02 b0 01 56 89 e7 f6 46 bb 80 74 13 66 |...L...V...F..t.f|
00000160 6a 00 66 ff 74 08 06 53 6a 01 6a 10 89 e6 48 80 |j.f.t..Sj.j...H.|
00000170 cc 40 cd 13 89 fc 5e c3 20 20 a0 0a 44 65 66 61 |.@...^...Defa|
00000180 75 6c 74 3a a0 0d 8a 00 05 0f 01 06 07 0b 0c 0e |ult:.....?|
00000190 83 a5 a6 a9 0d 0c 0b 0a 09 08 0a 0e 11 10 01 3f |.....|
000001a0 bf 44 4f d3 4c 69 6e 75 f8 46 72 65 65 42 53 c4 |.D0.Linu.FreeBS.|
000001b0 66 bb 44 72 69 76 65 20 00 00 80 8f b6 00 00 00 |f.Drive .....|
000001c0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 |.....|
*
000001f0 00 00 00 00 00 00 00 00 00 00 00 00 55 aa |.....U..|
00000200
```

* <https://en.wikipedia.org/wiki/Executable>

Basic Software Development Concept

- Source Code to Binary
 - Source Code → Compiler → Binary
 - ไม่สามารถแปลง Binary กลับมาเป็น Sour Code ที่สมบูรณ์ได้
- Source Code ถือเป็นความลับทางการค้า
- โปรแกรมส่วนมากไม่เปิดเผย Source Code
 - ผู้ใช้มีสิทธิ์ในการใช้โปรแกรม
 - ผู้ใช้ไม่มีสิทธิ์แก้ไข
 - การปรับปรุงแก้ไขเป็นของผู้ผลิตเพียงฝ่ายเดียว

Open Source Software

- ผู้ผลิตยินดีจะเปิดเผย Source Code
- เน้นเสรีภาพในการใช้ซอฟต์แวร์
- ผู้ใช้มีเสรีภาพในการ (ใช้ / แจกจ่าย / แก้ไข / ขาย) โดยไม่ต้องขออนุญาต
 - แต่ต้องคงสิทธิ์นี้ให้กับคนอื่นด้วยเช่นกัน
- เปิดโอกาสให้เกิดการช่วยกันพัฒนา
- Open Source ไม่ได้แปลว่า “ฟรี”
 - สามารถหารายได้จาก Open Source ได้
 - เปลี่ยนจากการขายผลิตภัณฑ์ (product) มาเป็นขายบริการ (service) เช่น Google, Facebook
- Open Source ไม่ได้แปลว่า “ไม่มีลิขสิทธิ์”

Open Source is not Freeware

- Shareware
- Freeware
- Postcard ware
- Freemium
- Open Source

Popular Open Source in Market

- Linux (Debian / Ubuntu / Redhat / ...)
- Server (Apache HTTPD Server / Nginx / MySQL / PostgreSQL)
- Google Chrome / Mozilla Firefox
- Wikipedia / WordPress / Joomla / Drupal
- VLC Media Player / Audacity / Freemind / GIMP
- OpenOffice / LibreOffice / Mozilla Thunderbird / 7-Zip / GnuCash
- C / C++ / C# / Java / JavaScript / PHP / Python / Dart / Flutter
- Visual Studio Code / Notepad++ / Netbeans / Eclipse

Understand in Software Licenses

- License = สัญญา “อนุญาต” ให้
 - ใช้งาน
 - พัฒนา/ปรับปรุง/แก้ไข
 - แจกจ่าย / เผยแพร่ / ขาย
- Open Source Licenses
 - สามารถ ใช้งาน / พัฒนา / ปรับปรุง / แก้ไข / ขาย ได้อย่างอิสระตามที่ให้สิทธิ์ไว้
- ผู้เขียนซอฟต์แวร์เป็นเจ้าของลิขสิทธิ์เสมอ

Popular Open Source Licenses

- Apache / BSD / MIT
 - นำ Source Code ไปใช้งานอย่างไรก็ได้ แต่ไม่มีสิทธิในการฟ้องร้องผู้พัฒนา
 - ต้องแนบ License File ไปด้วยเสมอ
- Creative Commons (CC)
 - CC-BY – ต้องอ้างอิงแหล่งที่มา
 - CC-SA – ให้ผู้นำไปใช้ต้องใช้ CC-SA เช่นกัน
 - CC-NC – ห้ามนำไปใช้เพื่อการค้า
 - CC-ND – ห้ามดัดแปลงแก้ไขใด ๆ
- GNU GPL – *“I open sourced my code so you should too”*
 - งานที่พัฒนาต่อ (derivative) ต้องใช้ GPL License และต้องแนบ License File ไปด้วยเสมอ
 - หากนำ Source Code ไปใช้ในงาน งานนั้นจะต้องใช้ GPL License ด้วย
 - ต้องเผยแพร่ Source Code หากมีการร้องขอ
- GNU LGPL
 - สามารถนำ Source Code ไปใช้ในงานได้
 - งานที่พัฒนาต่อ (derivative) ต้องใช้ GPL หรือ LGPL License และต้องแนบ License File ไปด้วยเสมอด้วย

What are the software licenses

- Python → Python Software Foundation license (PSF license)
- NumPy → BSD license
- OpenCV → Apache license
- pandas → New BSD license
- scikit-image → BSD license
- Python Imaging Library (PIL) → PSF license
- Matplotlib → PSF license

Thank You