

How Ai Works

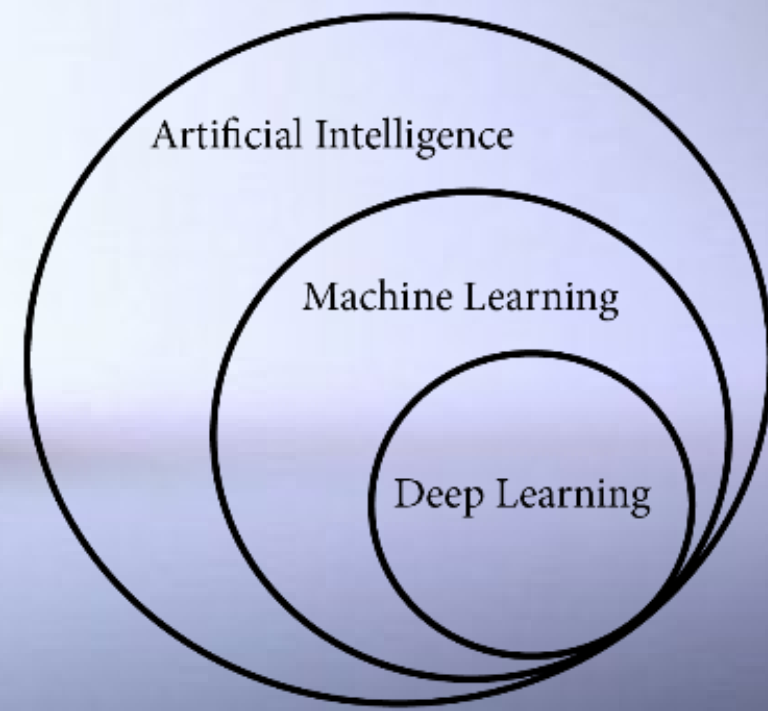
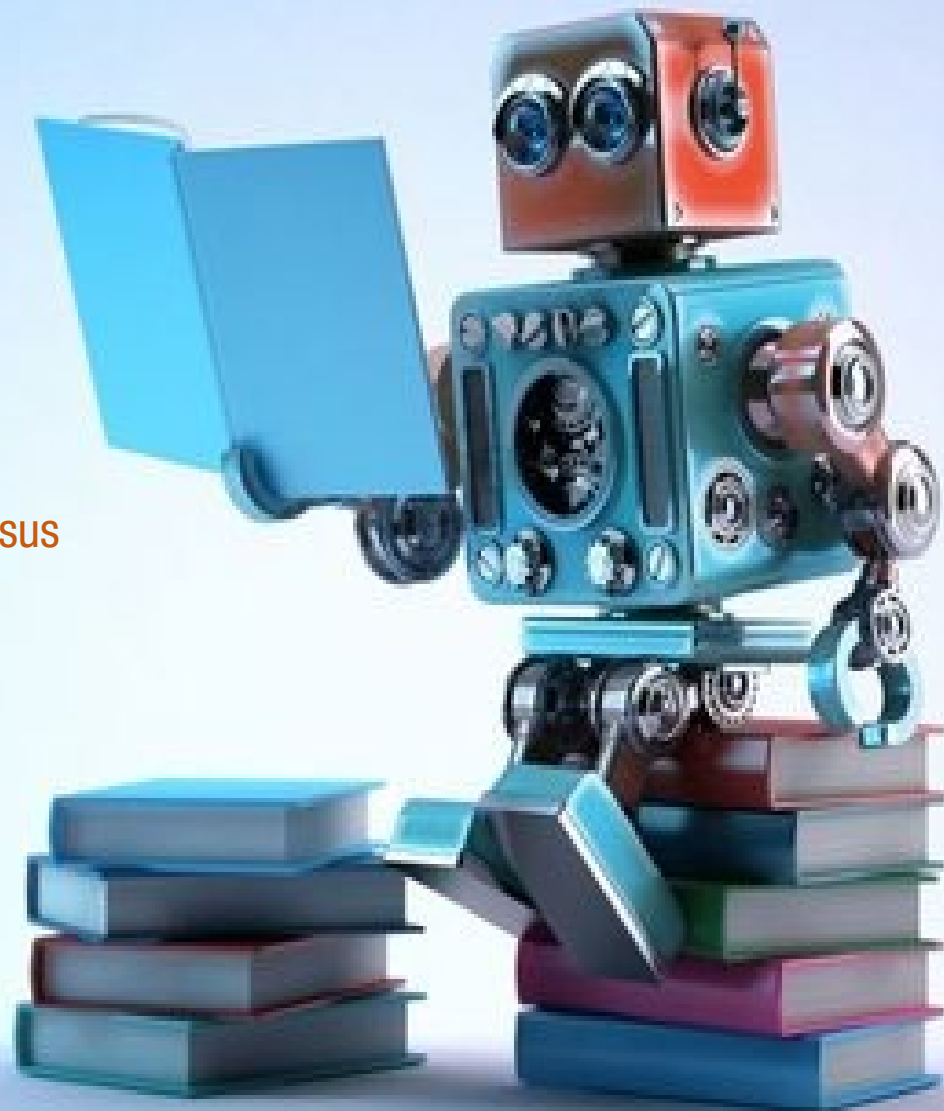
useful basics for librarians and researchers

Wayne de Fremery
Dominican University of California

AI

overview

General versus Narrow



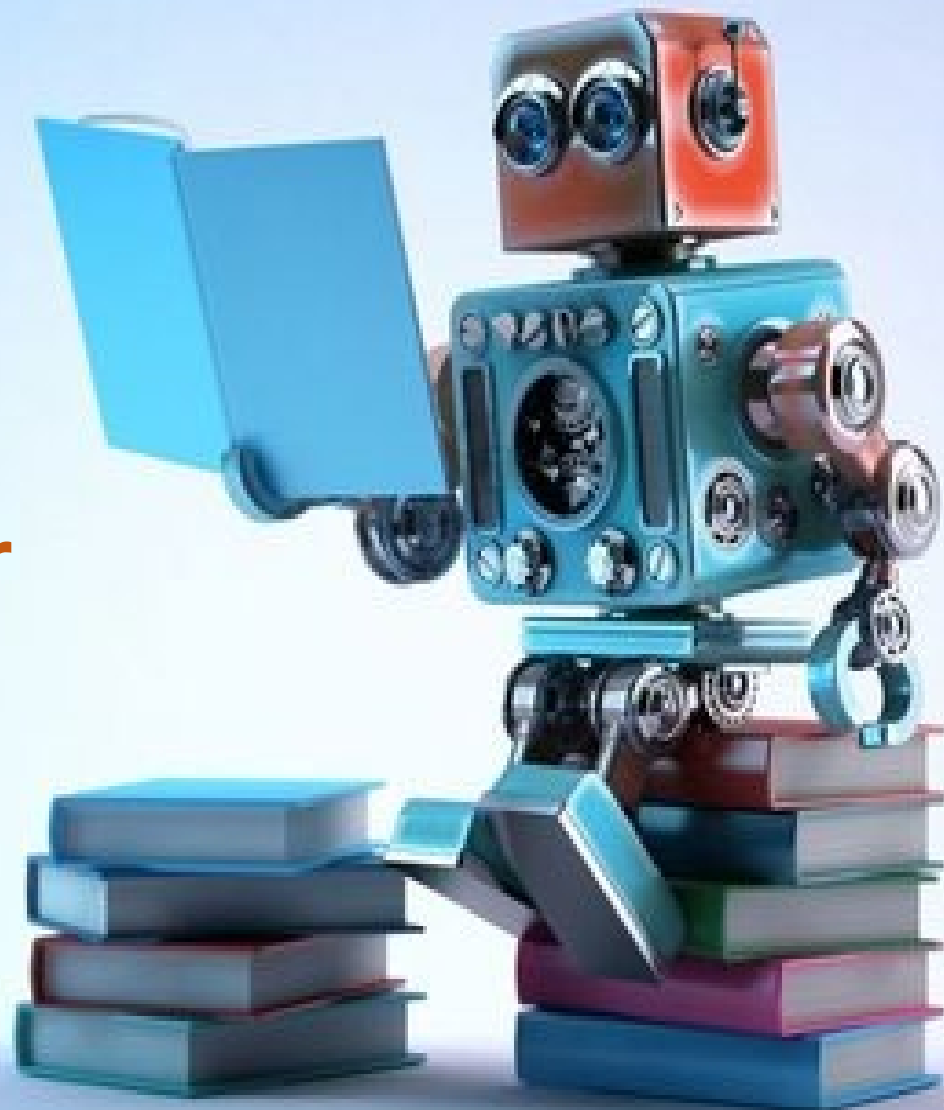
Kinds of Data
& Tasks

/

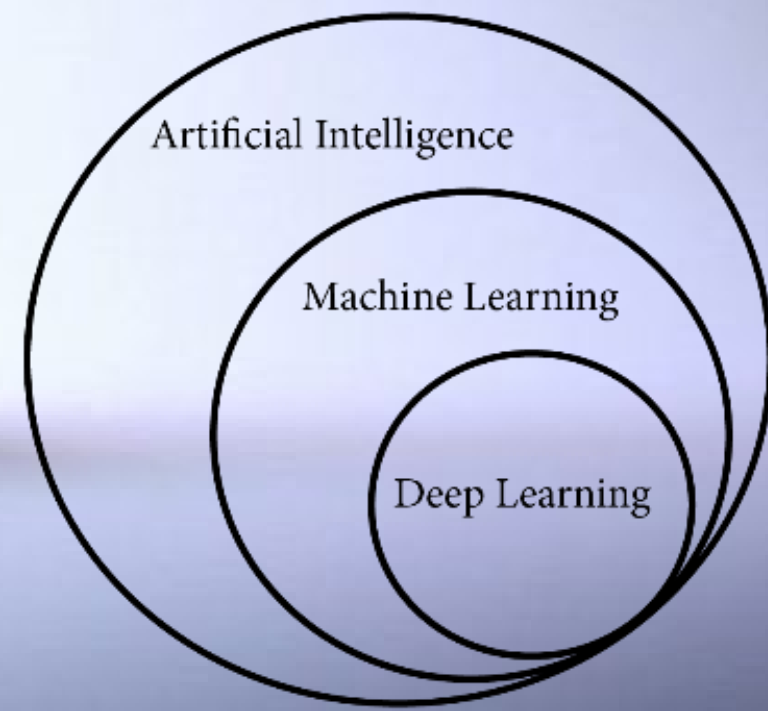
Kinds of Algorithms

AI & Deep Learning

Computer Vision

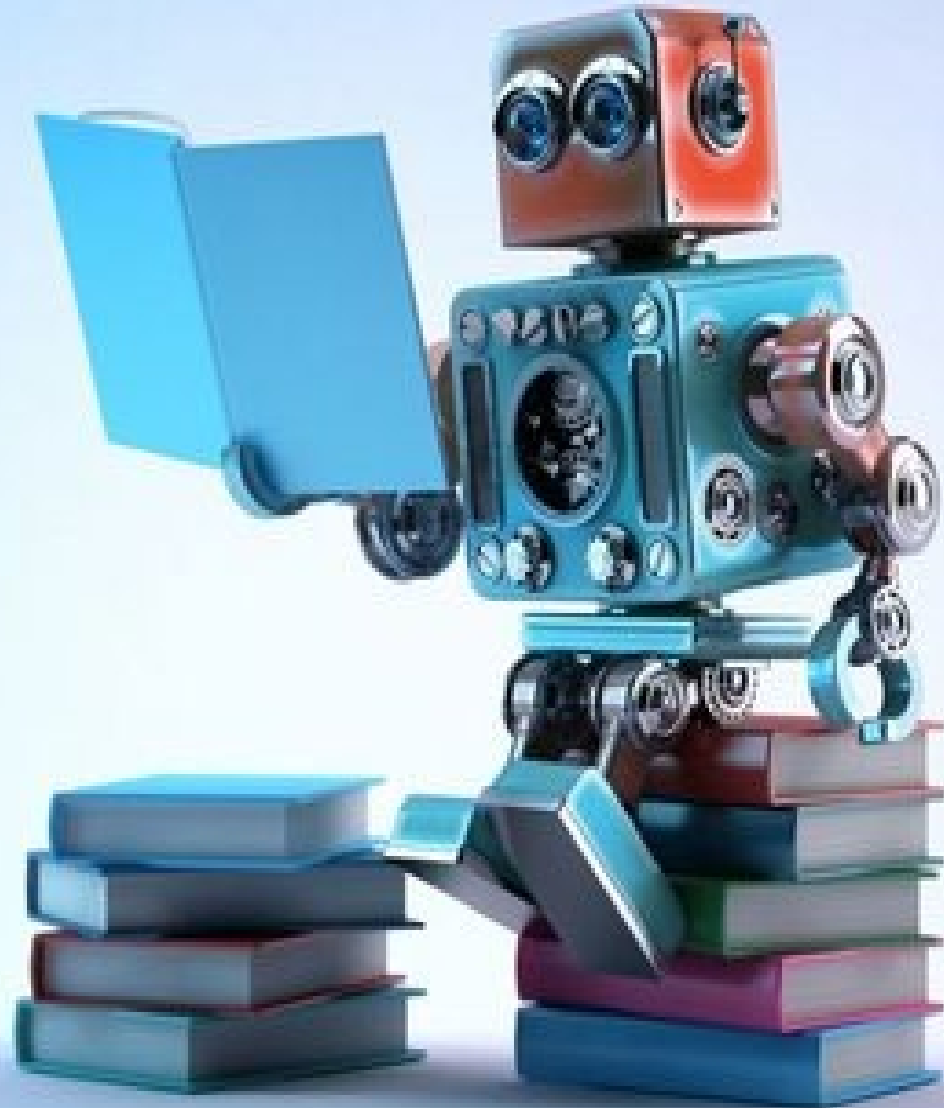


CNN (convolutional neural network)

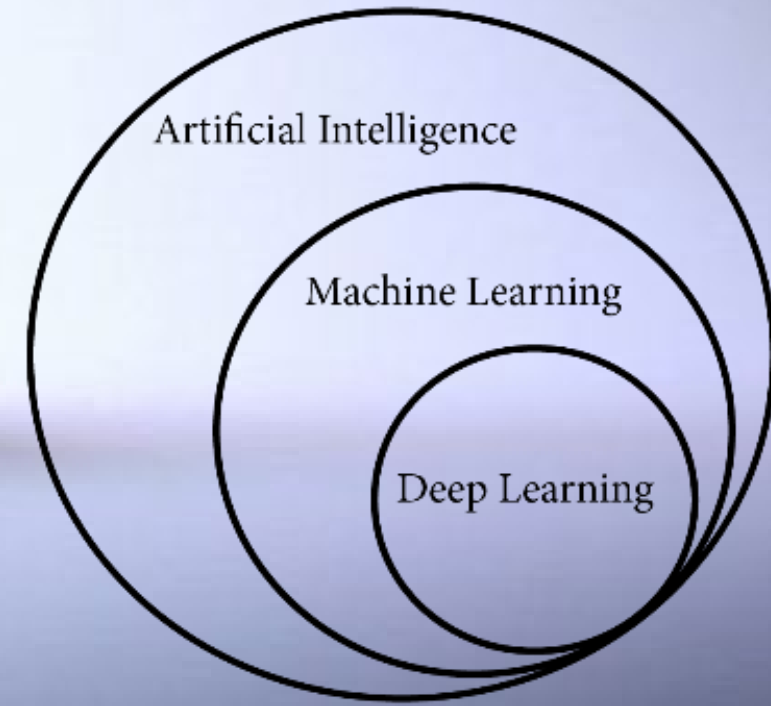


AI & Deep Learning

(N)atural (L)anguage (P)rocessing



RNN (recurrent neural networks)
BERT (Bidirectional Encoder Representations from Transformers)
GPT (Generative Pretrained Transformers)

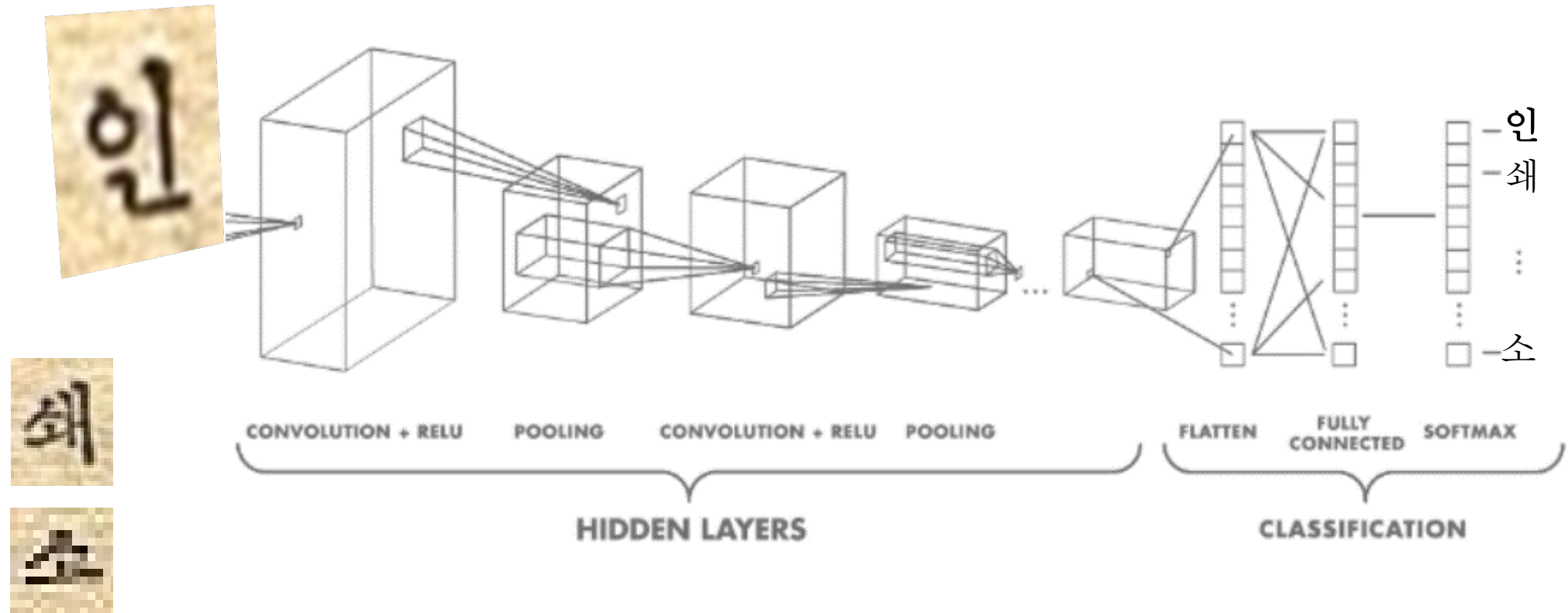


AI

Optical Character Recognition (OCR)

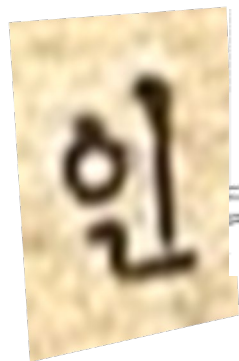
(using computer vision)

Deep Learning



convolutional neural network

Convolution



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

Input

plus

1	0	1
0	1	0
1	0	1

Filter / Kernel

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

1	1x1	1x0	0x1	0
0	1x0	1x1	1x0	0
0	0x1	1x0	1x1	1
0	0	1	1	0
0	1	1	0	0

4	3	

1	1	1x1	0x0	0x1
0	1	1x0	1x1	0x0
0	0	1x1	1x0	1x1
0	0	1	1	0
0	1	1	0	0

4	3	4

1	1	1	0	0
0x1	1x0	1x1	1	0
0x0	0x1	1x0	1	1
0x1	0x0	1x1	1	0
0	1	1	0	0

4	3	4
2		

1	1	1	0	0
0	1x1	1x0	1x1	0
0	0x0	1x1	1x0	1
0	0x1	1x0	1x1	0
0	1	1	0	0

4	3	4
2	4	

1	1	1	0	0
0	1	1x1	1x0	0x1
0	0	1x0	1x1	1x0
0	0	1x1	1x0	0x1
0	1	1	0	0

4	3	4
2	4	3

1	1	1	0	0
0	1	1	1	0
0x1	0x0	1x1	1	1
0x0	0x1	1x0	1	0
0x1	1x0	1x1	0	0

4	3	4
2	4	3
2		

1	1	1	0	0
0	1	1	1	0
0	0x1	1x0	1x1	1
0	0x0	1x1	1x0	0
0	1x1	1x0	0x1	0

4	3	4
2	4	3
2	3	

1	1	1	0	0
0	1	1	1	0
0	0	1x1	1x0	1x1
0	0	1x0	1x1	0x0
0	1	1x1	0x0	0x1

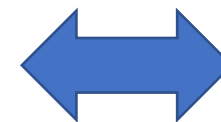
4	3	4
2	4	3
2	3	4



Descriptive features

4	3	4
2	4	3
2	3	4

Feature (map)



target

인

Text (Character) Encoding

Morse Code

(American, 1844; Continental, 1848; International, 1865)

	American (Morse)	Continental (Gerke)	International (ITU)
A	· · —	· · —	· · —
Ä	· · — · ·	· · — · ·	· · — · ·
B	· · · · —	· · · · —	· · · · —
C	· · — · ·	· · — · ·	· · — · ·
CH	· · — · ·	· · — · ·	· · — · ·
D	· · · · —	· · · · —	· · · · —
E	· · · ·	· · · ·	· · · ·
F	· · · · —	· · · · —	· · · · —
G	· · · · —	· · · · —	· · · · —
H	· · · · —	· · · · —	· · · · —
I	· · · ·	· · · ·	· · · ·
J	· · — · ·	· · — · ·	· · — · ·
K	· · — · ·	· · — · ·	· · — · ·
L	· · — · ·	· · — · ·	· · — · ·
M	· · — · ·	· · — · ·	· · — · ·
N	· · — · ·	· · — · ·	· · — · ·
O	· · — · ·	· · — · ·	· · — · ·
Ö	· · — · ·	· · — · ·	· · — · ·
P	· · — · ·	· · — · ·	· · — · ·
Q	· · — · ·	· · — · ·	· · — · ·
R	· · — · ·	· · — · ·	· · — · ·
S	· · — · ·	· · — · ·	· · — · ·
T	· · — · ·	· · — · ·	· · — · ·
U	· · — · ·	· · — · ·	· · — · ·
Ü	· · — · ·	· · — · ·	· · — · ·
V	· · — · ·	· · — · ·	· · — · ·
W	· · — · ·	· · — · ·	· · — · ·
X	· · — · ·	· · — · ·	· · — · ·
Y	· · — · ·	· · — · ·	· · — · ·
Z	· · — · ·	· · — · ·	· · — · ·
1	· · — · ·	· · — · ·	· · — · ·
2	· · — · ·	· · — · ·	· · — · ·
3	· · — · ·	· · — · ·	· · — · ·
4	· · — · ·	· · — · ·	· · — · ·
5	· · — · ·	· · — · ·	· · — · ·
6	· · — · ·	· · — · ·	· · — · ·
7	· · — · ·	· · — · ·	· · — · ·
8	· · — · ·	· · — · ·	· · — · ·
9	· · — · ·	· · — · ·	· · — · ·
0	· · — · ·	· · — · ·	· · — · ·
0 (alt)	· · — · ·	· · — · ·	· · — · ·

Bits					0	0	0	0	1	1	1	1				
					0	0	1	1	0	0	1	1				
b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	Column	Row	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p		
0	0	0	1	1	1	1	SOH	DC1	!	1	A	Q	a	q		
0	0	1	0	2	2	2	STX	DC2	"	2	B	R	b	r		
0	0	1	1	3	3	3	ETX	DC3	#	3	C	S	c	s		
0	1	0	0	4	4	4	EOT	DC4	\$	4	D	T	d	t		
0	1	0	1	5	5	5	ENQ	NAK	%	5	E	U	e	u		
0	1	1	0	6	6	6	ACK	SYN	&	6	F	V	f	v		
0	1	1	1	7	7	7	BEL	ETB	'	7	G	W	g	w		
1	0	0	0	8	8	8	BS	CAN	(8	H	X	h	x		
1	0	0	1	9	9	9	HT	EM)	9	I	Y	i	y		
1	0	1	0	10	10	10	LF	SUB	*	:	J	Z	j	z		
1	0	1	1	11	11	11	VT	ESC	+	;	K	[k	{		
1	1	0	0	12	12	12	FF	FS	,	<	L	\	l			
1	1	0	1	13	13	13	CR	GS	—	=	M]	m	}		
1	1	1	0	14	14	14	SO	RS	.	>	N	^	n	~		
1	1	1	1	15	15	15	SI	US	/	?	O	_	o	DEL		

American Standard Code for Information Interchange (1960s)

Text (Character) Encoding

American Standard Code for Information Interchange

Binary ↕	Oct ↕	Dec ↕	Hex	Abbreviation			Unicode Control Pictures ^[b]	Caret notation ^[c]	C escape sequence ^[d]	Name (1967) ↕
				1963 ↕	1965 ↕	1967 ↕				
000 0000	000	0	00	NULL	NUL		NUL	^@	\0	Null
000 0001	001	1	01	SOM	SOH		SOH	^A		Start of Heading
000 0010	002	2	02	EOA	STX		STX	^B		Start of Text
000 0011	003	3	03	EOM	ETX		ETX	^C		End of Text
000 0100	004	4	04		EOT		EOT	^D		End of Transmission
000 1010	012	10	0A		LF		LF	^J	\n	Line Feed
000 1011	013	11	0B	VTAB	VT		VT	^K	\v	Vertical Tab
000 1100	014	12	0C		FF		FF	^L	\f	Form Feed
000 1101	015	13	0D		CR		CR	^M	\r	Carriage Return ^[h]

10 = line feed

13 = carriage return

32 = space

48 = 0

65 = A

? = 인

Binary ↕	Oct ↕	Dec ↕	Hex	Glyph		
				1963 ↕	1965 ↕	1967 ↕
010 0000	040	32	20	space		
011 0000	060	48	30	0		
100 0001	101	65	41	A		

Unicode

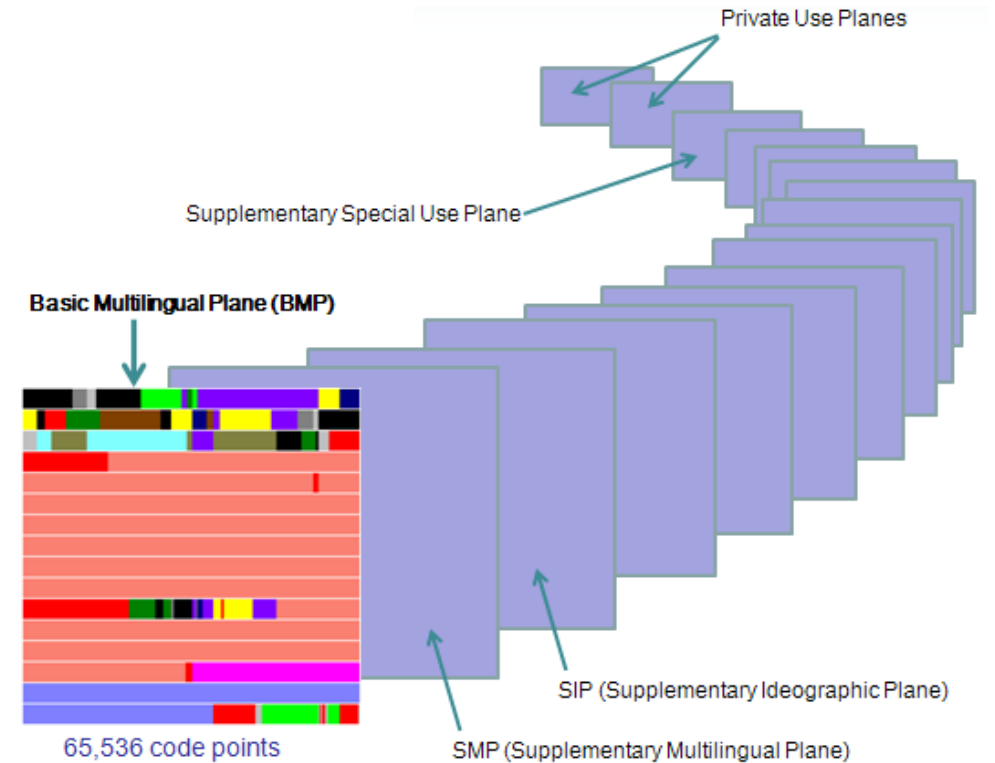
Unicode Consortium 501 (C)3 Mountain View CA

A universal character encoding scheme
initiated in 1987. The Unicode Consortium was
incorporated in California 1991.

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

- Latin scripts and symbols
- Linguistic scripts
- Other European scripts
- African scripts
- Middle Eastern and Southwest Asian scripts
- Central Asian scripts
- South Asian scripts
- Southeast Asian scripts
- Hangul
- CJK Unified Ideographs
- American scripts
- Symbols
- Diacritics
- UTF-16 surrogates and private use
- Miscellaneous characters
- Unallocated code points

Unicode Code Spaces



Sources:

(above) http://upload.wikimedia.org/wikipedia/commons/8/8e/Roadmap_to_Unicode_BMP.svg (modified by author);
(right) <http://www.w3.org/International/articles/definitions-characters/>

Digital Fonts

Glyphs
mapped to
encoding

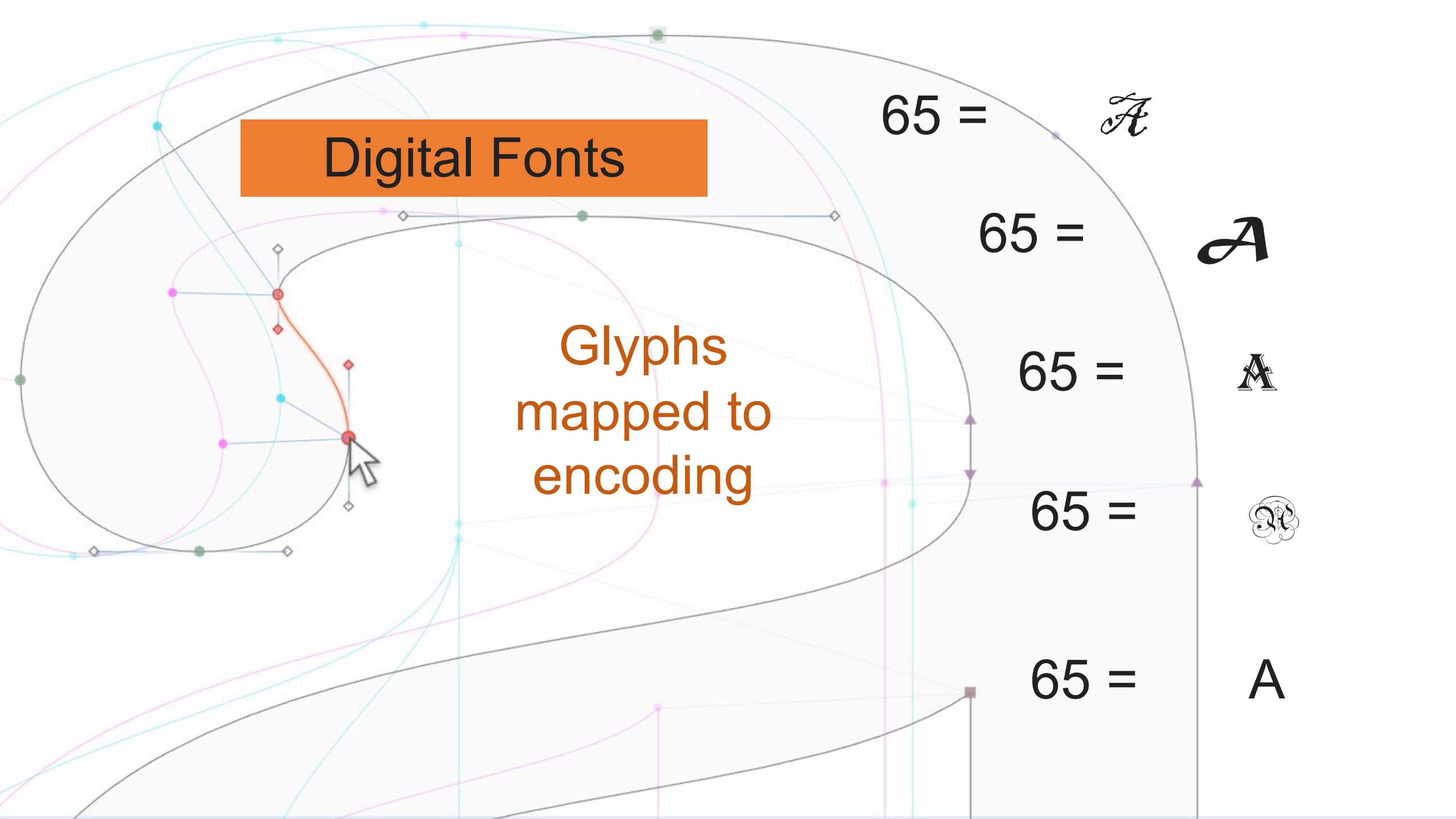
65 = *A*

65 = **A**

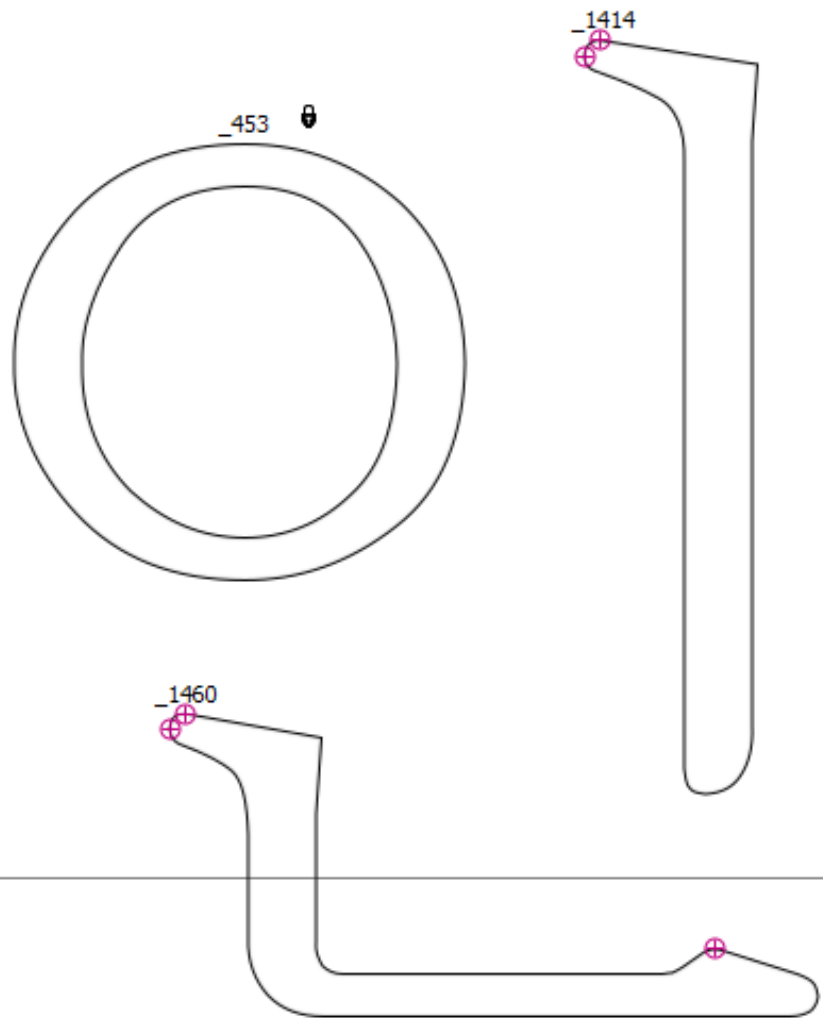
65 = **A**

65 = **A**

65 = **A**



Digital Fonts



236 157 184 = 인

236 157 184 = 인

236 157 184 = **인**

236 157 184 = 인

Glyphs
mapped to
encoding

Text Encoding

진달내곶

나보기가 역겨워
가실재에는
말업시 고히 보내드리우리다

寧邊에藥山
진달내곶
아름싸다 가실길에 썩리우리다

가시는거름거름
노힌그곶출
삼분히즈려밧고 가시웁소서

나보기가 역겨워
가실재에는
죽어도아니 눈물흘니우리다

진달내□

나보기가 역겨워
가실□에는
말업시 고히 보내드리우리다

寧邊에藥山
진달내□
아름□다 가실길에 □리우리다

가시는거름거름
노힌그□출
삼분히즈려밧고 가시웁소서

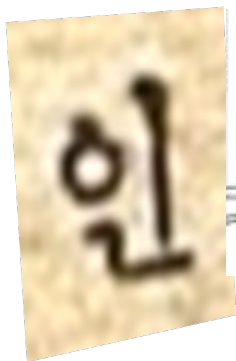
나보기가 역겨워
가실□에는
죽어도아니 눈물흘니우리다

Haansoft Batang

한글
폰트

51652458044523659882131013104
52084837244592440323250669442
00508921310440324989260031506
40457161310475685062949884324
42565517632483724523646300475
32508644753245796131013106383
93700250640342772366532131051
65245804452365988213105050047
49260017457963244032498924460
05064032602624753250864475324
57961310131044032498844571644
14447492441444749213104543255
18044536598825277613104934148
51655176515924714048159442563
24403249884507414954849436131
01310452084837244592440323250
66944200508921310440324989260
03150640457161310514535061246
02050500457683245576479325512
8457685086447532457960

Convolution



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

Input

plus

1	0	1
0	1	0
1	0	1

Filter / Kernel

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

1	1x1	1x0	0x1	0
0	1x0	1x1	1x0	0
0	0x1	1x0	1x1	1
0	0	1	1	0
0	1	1	0	0

4	3	

1	1	1x1	0x0	0x1
0	1	1x0	1x1	0x0
0	0	1x1	1x0	1x1
0	0	1	1	0
0	1	1	0	0

4	3	4

1	1	1	0	0
0x1	1x0	1x1	1	0
0x0	0x1	1x0	1	1
0x1	0x0	1x1	1	0
0	1	1	0	0

4	3	4
2		

1	1	1	0	0
0	1x1	1x0	1x1	0
0	0x0	1x1	1x0	1
0	0x1	1x0	1x1	0
0	1	1	0	0

4	3	4
2	4	

1	1	1	0	0
0	1	1x1	1x0	0x1
0	0	1x0	1x1	1x0
0	0	1x1	1x0	0x1
0	1	1	0	0

4	3	4
2	4	3

1	1	1	0	0
0	1	1	1	0
0x1	0x0	1x1	1	1
0x0	0x1	1x0	1	0
0x1	1x0	1x1	0	0

4	3	4
2	4	3
2		

1	1	1	0	0
0	1	1	1	0
0	0x1	1x0	1x1	1
0	0x0	1x1	1x0	0
0	1x1	1x0	0x1	0

4	3	4
2	4	3
2	3	

1	1	1	0	0
0	1	1	1	0
0	0	1x1	1x0	1x1
0	0	1x0	1x1	0x0
0	1	1x1	0x0	0x1

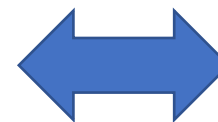
4	3	4
2	4	3
2	3	4



Descriptive features

4	3	4
2	4	3
2	3	4

Feature (map)

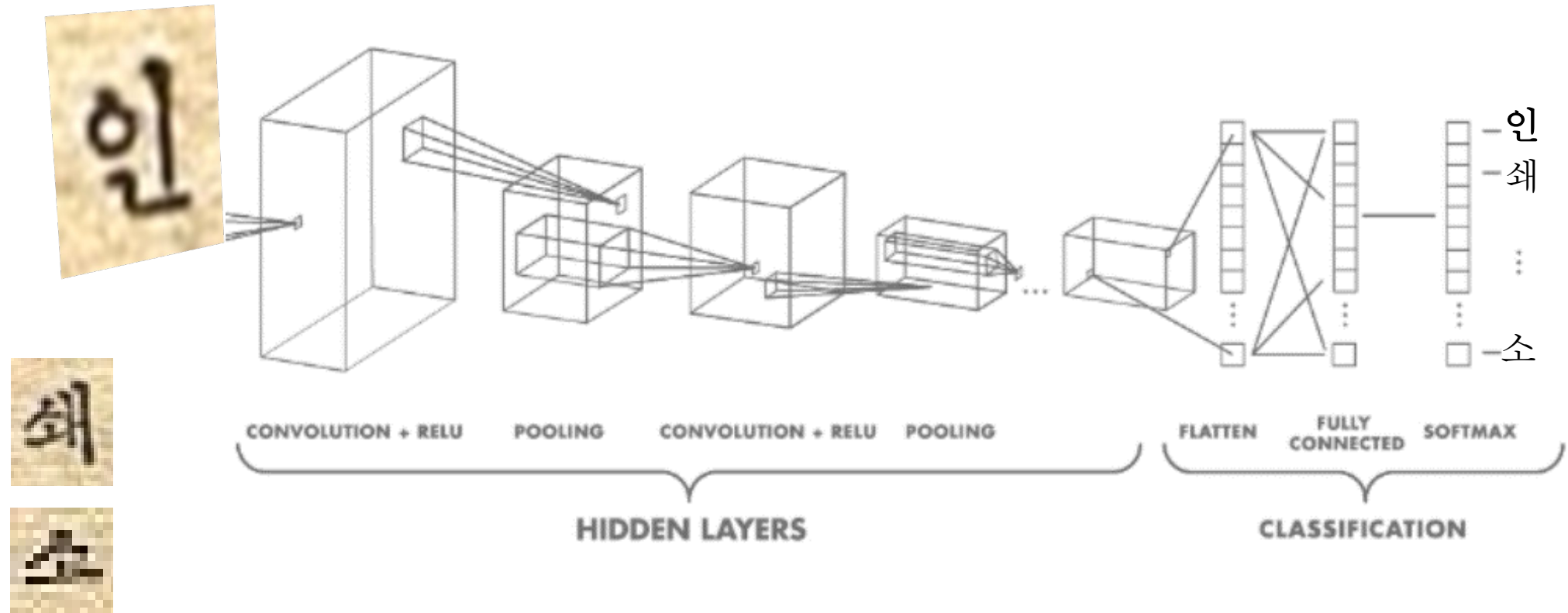


target

236 157 184

(인)

Deep Learning



convolutional neural network

CNNs & Inference

가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각
각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각
각	각	각	각	각	각	각	간	간	간	간	간	간	간	간	간



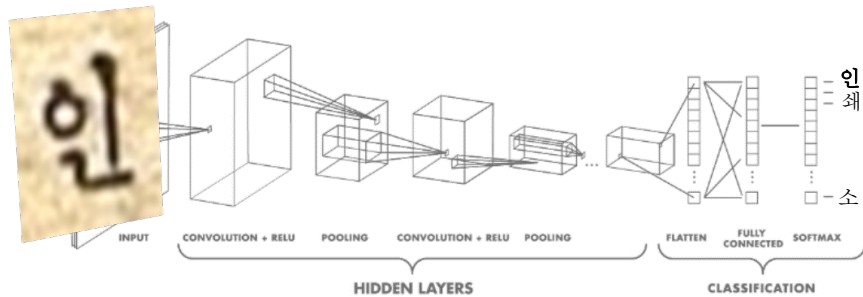
Weight File

(record of which features best predict/ describe objects in training set)

Something.weights

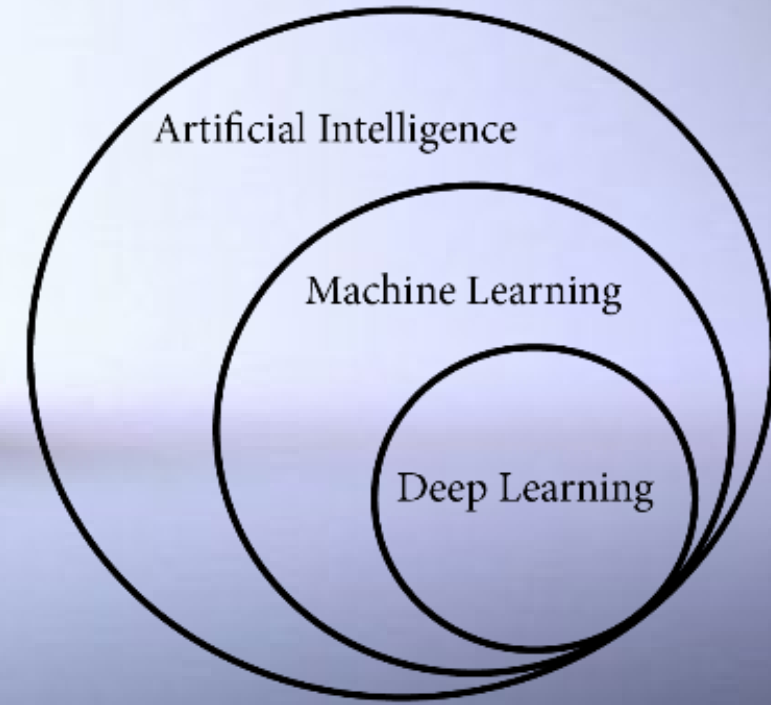
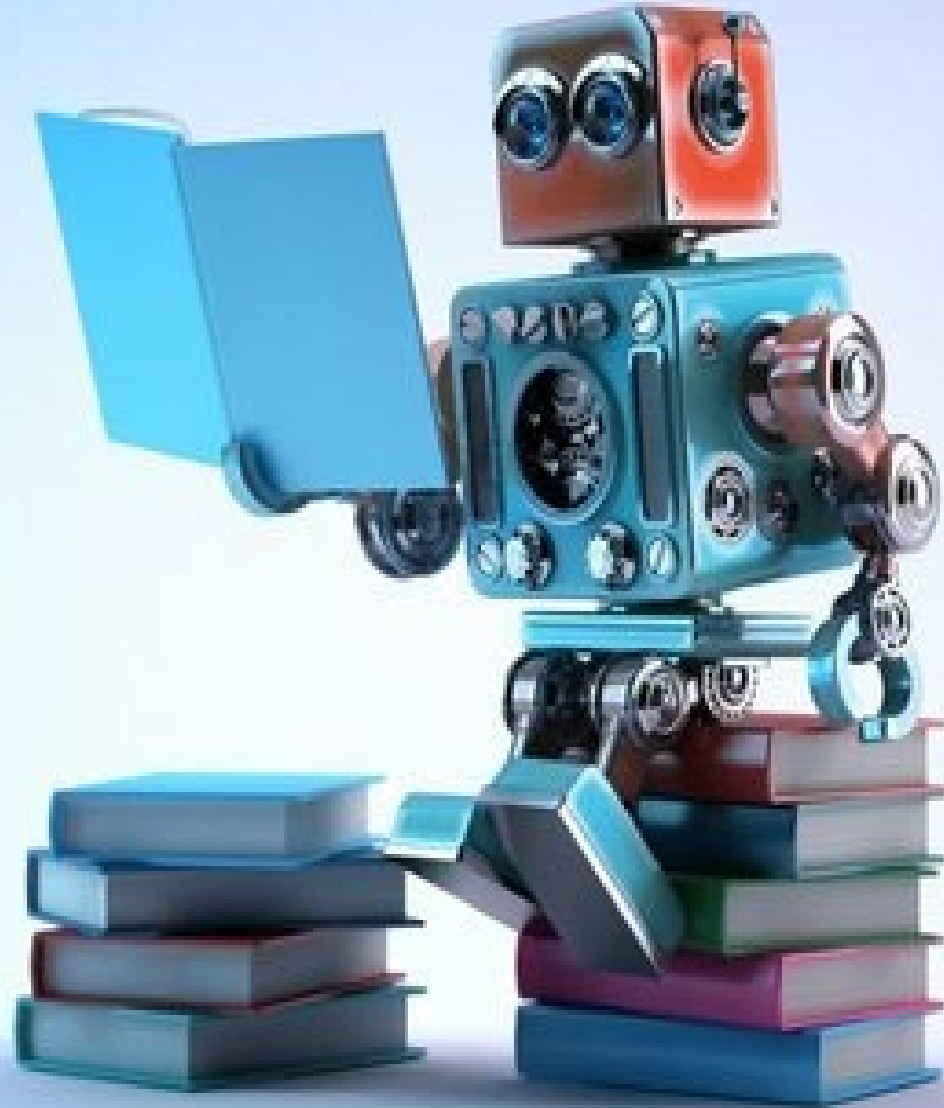


236 157 184



AI & Deep Learning

(N)atural
(L)anguage
(P)rocessing



(G)enerative (P)retrained (T)ransformer

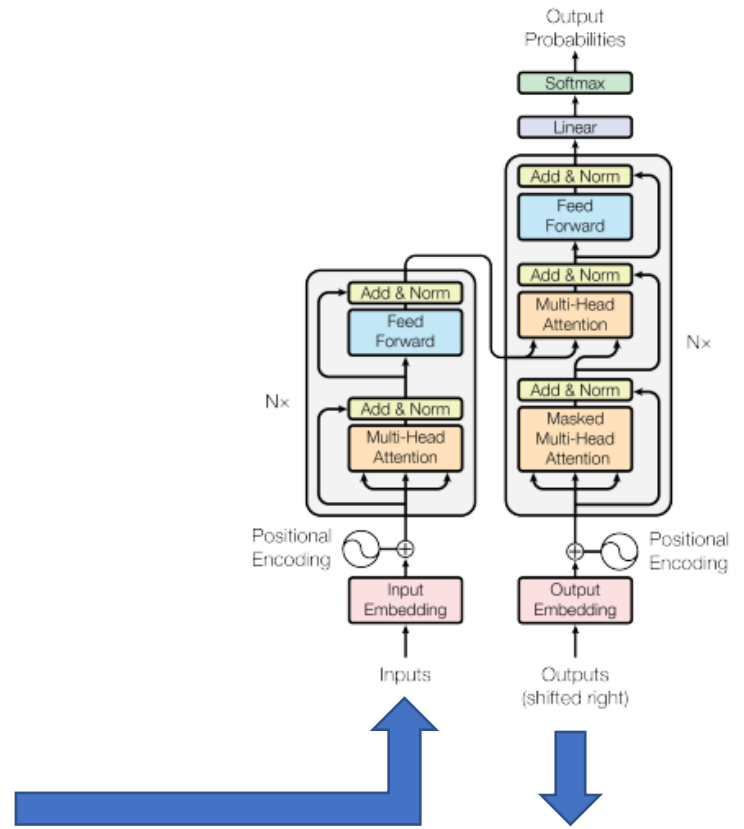
```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<topics>
<topic>
<perio>로동신문</perio>
<date>1950년 6월 25일</date>
<side>3</side>
<title></title>
<part>보도기사</part>
<field></field>
<region></region>
<title>철도수송력의 강화 위해 기관차실동률 제고에!</title>
<stitle>신북청기관구 로동자들</stitle>
<text>
```

철도수송력의 강화위해 기관차실동률 제고에 ! 신북청기관구 로동자들 신북청기관구 로동자들은 8.15해방 5주년을 철도수송력의강화로써 뜻깊게 맞이하고자 기관차 수리의 질적제고를 위한 투쟁을 가강히 전개하고있다. 그들은 우선 기관차의 질적수리제고를 위하여 낡은것을 용감히 퇴치하고 새로운 작업방법을 채용하는데 창발적노력을 기울이고있다. 수리직장로동자들은 40여종에 달하는 수리공구들을 품종별로재검토한후 마모되고 일부수사되고그들을 전부 자체제작하여 새것으로 보충하였으며 수리작업의 중요부속품인 각종바네등 종전에는 철도수송력의강화를 위하여 3분선반 작기수리개조와 신문식동무의 6척선반 작기수리개조와 같은 일을 하고있다. 이 곳 로동자들은 사업과정에서 일상적으로 호상 관계를 긴밀히하며 상호협력하는 분위기를 조성하여 생산협의회 등의 역할을 높이며 성능이 좋치못한 기관차를 성능이 우수한기관차로 교체하고있다. 특히 이곳 로동자들은 5호 미가 44호등은 종전에는 자체성능을 70%밖에 발휘하였으나 현재는 정이 규격대로되어있지않다는 로동자들의개별적 의견을 중시하여 정시로 견인하고있다. 특히 이곳 로동자들은 지난 5월중에는 울동휴차로 인하여 재생 시킴에 성공하였으며 국가에 막대한 리익을 주었다. 그들은 푸렛치 4호 재생을 위하여 정기적으로 되는 생산협의회에서 이 의 재생에대한 구체적 방법들을 충분히 토의한후 매개 로동자들의 기능수준에 엄격히 기초하여 분공을 정확히 조직하였다. 이리하여 6월에 들어서면서부터 기관차의 실동률을 부단히 높이고있는바 현재 기대당 실동률은 1.4분기에 비하여 7%나 더 제고시키였다. 이곳 로동자들은 기관차수리의 질적 제고에 더한층 창발적 노력을 기울임으로써 연간주행계획을 자기들이 맹세한 10월 15일까지에 반드시 초과완수할 굳센 결의밑에 계속 힘차게 싸우고있다. 통신원 윤지월</text>

```
<author>통신원 윤지월</author>
</topic>
</topics>
```

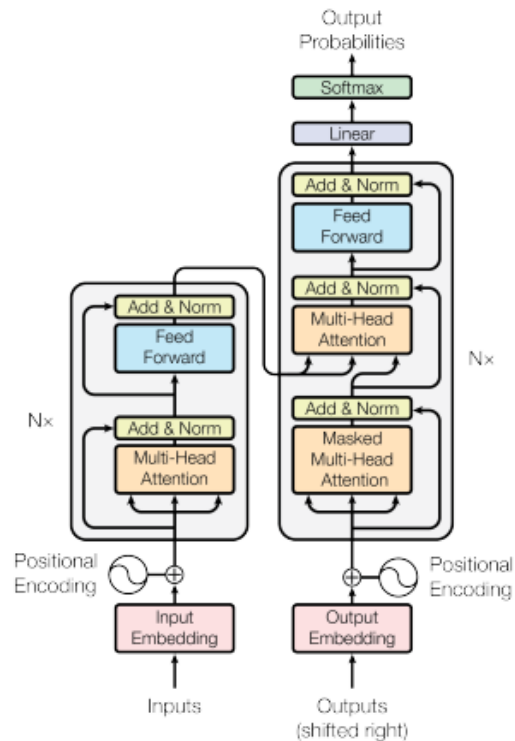
Vast amounts of historical, encoded text as input



```
<?xml version="1.0" encoding="UTF-8"? ...
```

<철도수송력의 강화위해 기관차실동률 제고에 ! ...

<?xml version="1.0" encoding="UTF-8"?>



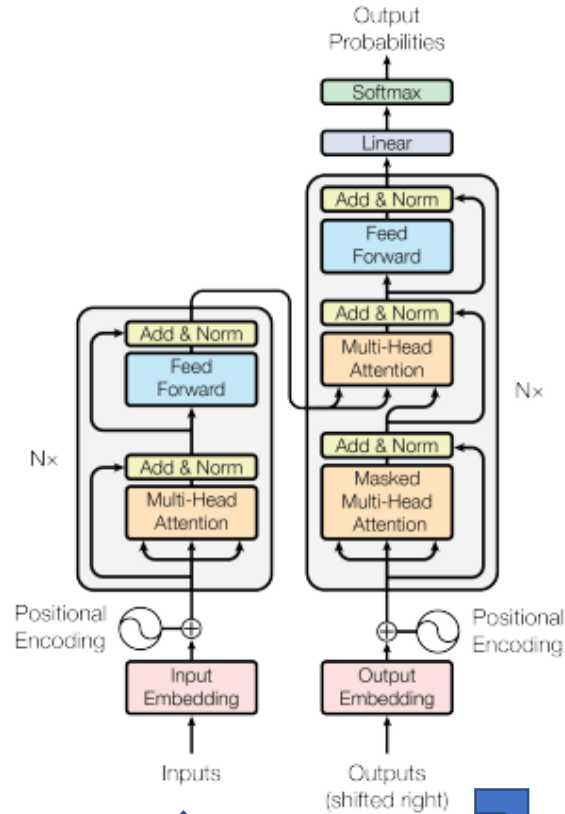
(G)enerative
(P)retrained
(T)ransformer

<?xml version



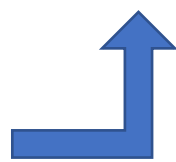
="1.0" encoding="UTF-8"?>

철도수송력의 강화위해 기관차실동률 제고에 !



(G)enerative
(P)retrained
(T)ransformer

철도수송력의



강화위해 기관차실동률 제고에 !

Optical Character Recognition (OCR)

Computer
Vision

Optical Character Recognition (OCR)

Computer Vision

(+)

Natural Language Processing

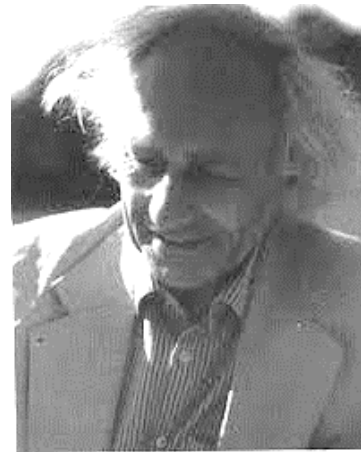
AI

a bibliographic process

copies &

accounts of difference

“Bibliography is the discipline that studies texts as recorded forms, and the processes of their transmission, including their production and reception.”



D. F. McKenzie. *Bibliography and the Sociology of Texts*

“The domain of information science is the transmission of the universe of human knowledge in recorded form, centering on manipulation (representation, organization, and retrieval) of information, rather than knowing information”

Association for Information Science and Technology website, “What is information science?” accessed February 3, 2022, <https://www.asist.org/about/what-is-information-science/>.



— Tefko Saracevic. (2009). Information science. In M.J. Bates (Ed.), *Encyclopedia of library and information sciences* (3rd ed.) (pp. 2570–2585). New York: Taylor and Francis.

Bibliography

Βιβλιογραφία book-writing

the writing out of books

書 글 서 誌 기록할 지 學 배울 학

writing-record

recording/ documenting writing / study of

Probably Japanese neologism. *Chosŏn ilbo* makes reference to a new Japanese journal called <<書誌學>> on January 25, 1933.

Bibliography

“List Makers”

“Studiers of Texts”

Enumerative

Descriptive

Analytical

Critical



Information Science



Humanities

Patrick Wilson *Two Kinds of Power: An Essay on Bibliographical Control* (1968)

Bibliography

Enumeration

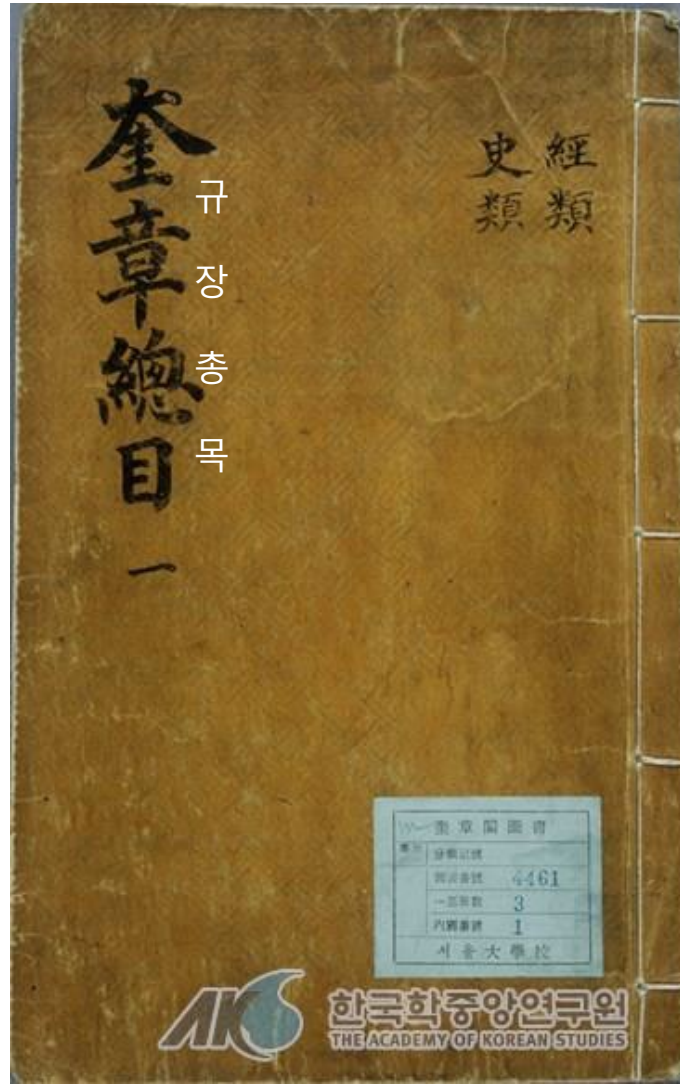
Description

Analysis

Critique

of

Copies



目錄學

Enumerative Bibliography



Descriptive Bibliography

How Poetry Mattered in 1920s Korea

1. *Onoe ūi mudo* (Dance of anguish)

Author(s):
Paul Verlaine,
Remy de Gourmont,
Albert Samain,
Charles Baudelaire,
William Butler Yeats, et al.

Translator:
Kim Ōk

(In the Hwabong
collection)



Dimensions (w x h in cm): 13.8 x 19.8; spine 7.4 (mm);
4.6-*p'anhyŏng*; proportions (width to height): 1:1.435
Cover materials: coated card (thickness: .34 mm)
Color(s): three (magenta, orange, green)
Image: flowers and musical staff

¹ Except when I discuss *Chindallaekkor* I have used the term "edition" as it is used in Korean bibliographic practice to suggest *p'anbon* 版本. Please see my discussion of "p'anbon," "edition," "issue," and "state" in Chapter Five.

Descriptive Bibliography



Onoe ū mudo title page

Onoe ū mudo Colophon

Onoe ū mudo

inswae (printing): March 15, 1921
parhaeng (release date): March 20, 1921

price: 1 wŏn

p'yŏnjip kyŏm/
parhaengja (editor/publisher):
Ko Kyŏng-sang

Kyŏngsŏng Chongno
2-chŏngnok 181

inswaein (printer):
Kim Sŏng-p'yo

Kyŏngsŏng Hwanggŭnyŏng
1-chŏngnok 181

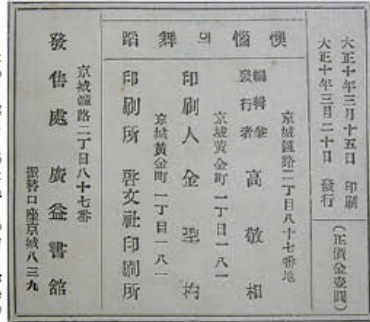
inswaeso (place of printing):
Kyeumusa Inswaeso

Kyŏngsŏng Hwanggŭnyŏng
1-chŏngnok 181

palsuch'ŏ
(publisher/distributor):
Kwangik Sŏgwan

Kyŏngsŏng Chongno
2-chŏngnok 87

chŭch'e kujwa (account
number to which money can be
sent): Kyŏngsŏng 839



Onoe ū mudo Bibliographic Notes

Endsheets Notes

_two different kinds.
a) .07 mm feels coated.
b) .03 mm tissue paper.

Title Page and Front Matter Notes

_title page:
single color.
3 stamps.

heavy stock (.12 mm).
relief printed, can feel
impressions of type.

_sheet with authors names:
heavy stock (.13 mm), but
does not feel coated.

_half title page:
same as body paper; feels
slightly coated.

Notes on Margins

-

Paper Notes

_appears that many different
kinds of paper were used: pp.
3-18 low opacity (.36 mm); pp.
19-34 more opaque (.37 mm);
pp. 35-end, different paper.

_no chain lines.
_glue holds pg. 18 to pg. 19 at
top and bottom.
_fibers clearly visible.

Notes on Typefaces

_pg. 3 title face 〇體 .5 cm x .5
cm; sam ㄱ, the first syllable in

Number of Pages—174; extra
sheets: (+1 front endsheet/ +1
tissue paper/ +1 title page/ +1
authors names/ +1 half title)

Paper (body)
1 sheet .04 mm
8 sheet .63 mm

Margins (in cm)

Page No.	Top	Bottom	Outside	Gutter
pg. 3	2.2	1.9 to folio	2 to dates	
pg. 6	2.6 to text; 4.4 to title	1.5 to folio; 2.3 to text	2 to title; 3.5 to text	1.7
pg. 15	3.4 to text; 4.5 to title	1.6 to folio	2.1	1.7

the body face is .3 cm x .3 cm.
_on pg. 15 ka ㄱ in title case is
.4 cm x .4 cm; ka ㄱ in first line
of body is .3 cm x .3 cm.

General Notes

_someone had fun "correcting"
the text with a red pen. On pg.
19 pul ㅍ is annotated with pur
ㅍ.

Binding Notes

_yangjang
_paper over boards; endsheets
glued to cover stock.

_sewn so that string is visible
between pp. 10-11, 26-27,
42-43, 58-59, 74-75, 90-91,
106-107, 122-123, 138-139,
154-155, 170-171.

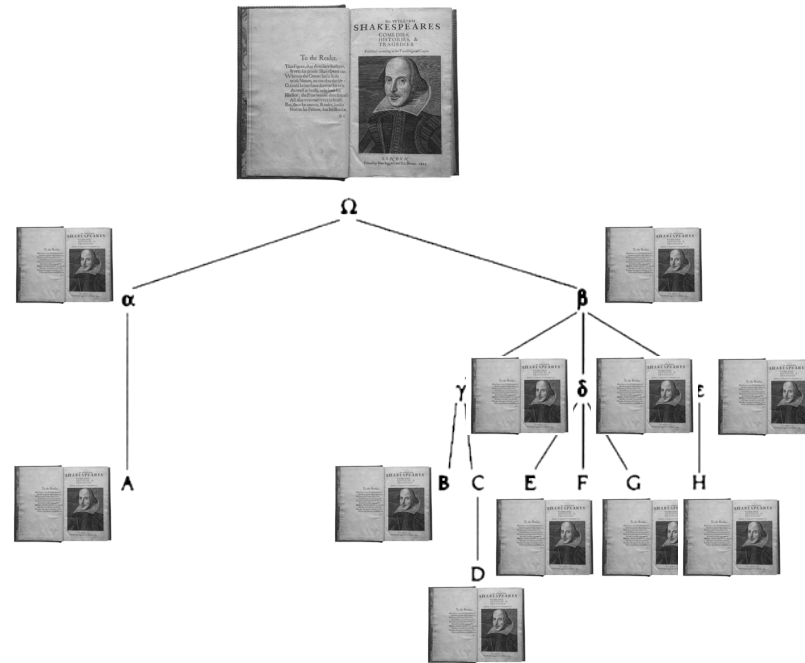
4.3	.
6.3	:
6.9	.
13.5	.
14	.



New Bibliography /
Analytical
Bibliography

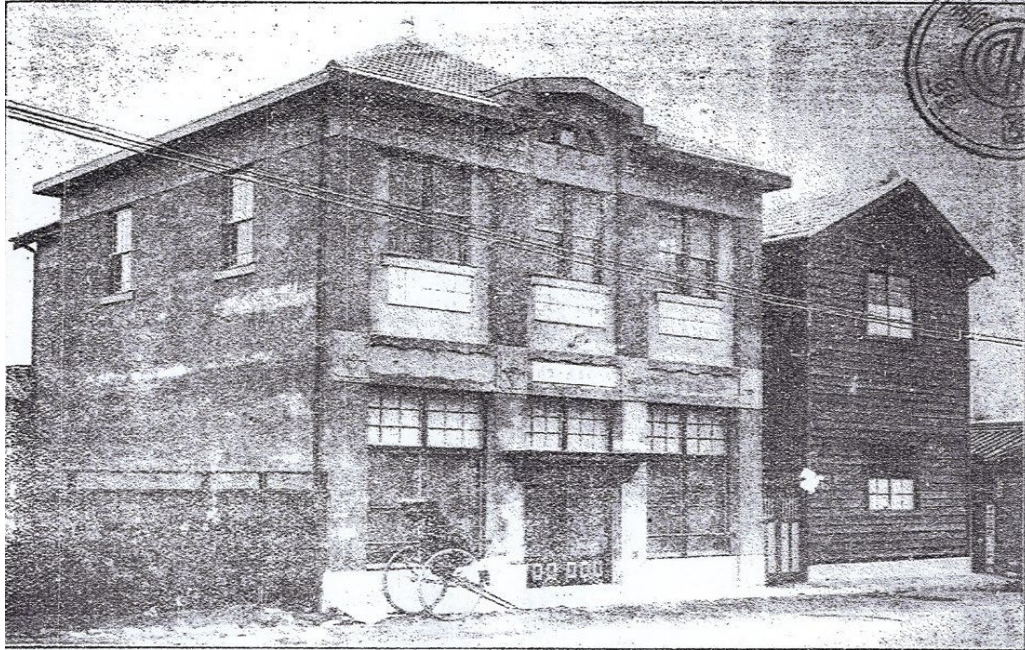
“Greg’s authority [was] ...
rooted in classical and
medievalist
methodologies of textual
criticism[; it] pivoted on
the organicist historicism
of stemmatics.”

The Cambridge Companion to Textual Scholarship
(Cambridge Companions to Literature) (Page 79).
Cambridge University Press. Kindle Edition.



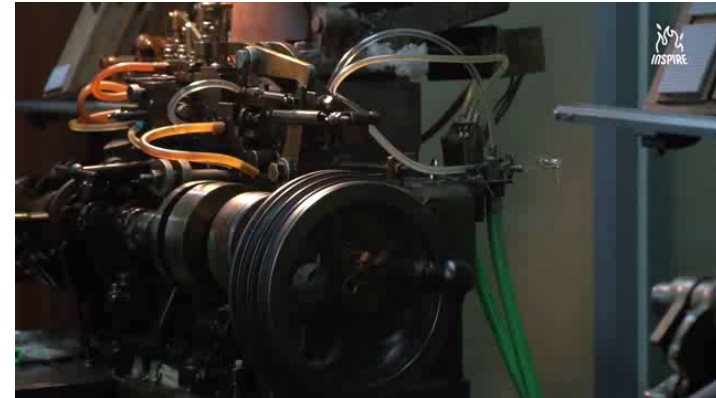
[Books] are products
of a particular
manufacturing or
technical process,
and the study of that
process for printed
texts is usually called
*analytical
bibliography.*”

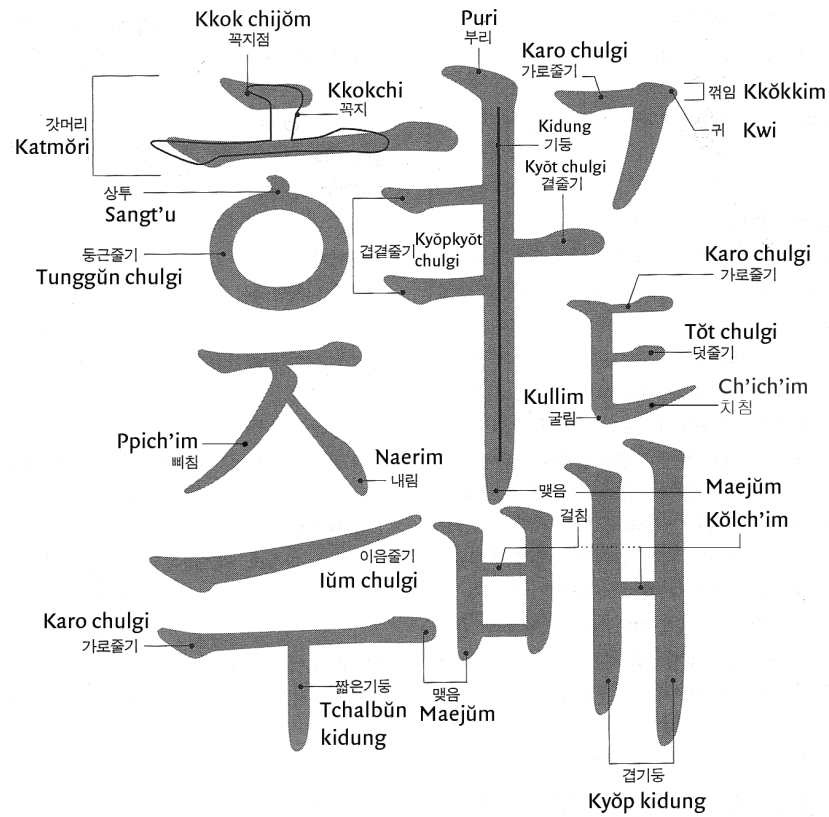
Greetham, David. *Textual Scholarship: An
Introduction* (Garland Reference Library
of the Humanities) (Kindle Locations 438-
441). Taylor and Francis. Kindle Edition.



Newly built Hansong Tosŏ Chusik Hoesa building, *Haksaenggye* (December 1920), unnumbered front matter (image from microfilm at the National Library of South Korea). The work of establishing Hansong Tosŏ's printing facility was completed in April of 1921 and the first materials were printed there in May by No Ki-jŏng. Before May of 1921, materials published by Hansong Tosŏ were printed at Chŏe Nam-sŏn's Sinmun'wan.

Analytical Bibliography





Source: Han Chae-jun 한재준, ed., *Han'gūl kälkkol yongō sajōn*
 한글글꼴용어사전 (A dictionary of typographic terms for han'gūl)
 (Seoul: Sejong Taewang Kinyōm Saōphoe, 2000), 46.

Critical Bibliography (Textual Studies/ Critical Editing)



The Covers of the Hansŏng Tosŏ issue (left) and the Chungang Sŏrim issue (right) in the Appenzeller-Noble Memorial Museum and the Museum of Contemporary Korean Poetry (Han'guk Hyŏndaesi Pangmulgwan), respectively.



Spine of the Hansŏng Tosŏ issue (left) and the Chungang Sŏrim issue (below).

Sources: Images of Hansŏng Tosŏ issue are from the Hwabong Mun'go (far left) and Appenzeller-Noble Memorial Museum collections (right). Image of the Chungang Sŏrim issue is from the collection of Ch'oe Ch'ŏr-hwan.

소래도업시

소래도업시

소리도업시

소리도업시

Soraŕsori 소래/소리 (sound) page 73 in the Chungang Sörim issue (left, in the collection of Ch'oe Ch'ör-hwan collection), Hansöng Tosö issue (center, in Öm Tong-söp collection) and in the Munhak Sasang *yöngi* nbon.





Table 5.2 Discrepancies between the Two 1925 Issues of *Chindallaekkot* and Important Collected Works of Kim So-wŏl

No.	Page No.	Hansŏng Tosŏ Issue	Chungang Sŏrim Issue	Munhak Sasang <i>Yonginbon</i>	Hanyang University <i>Yonginbon</i>	Kwŏn Yŏng-min*	Kim Chong-uk†	Kim Yong-jik‡	O Ha-gŭn§	Cho Tong-il and Yun Chu-ŭm**
1	73	In the last line of the poem, the word “sound” is printed “ <i>sorae</i> 소래.”	In the last line of the poem, the word “sound” is printed “ <i>sorae</i> 소래.”	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리.”	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리.”	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리” (pg. 132).	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리” (pg. 222).	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리” (pg. 73).	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리” (pg. 83).	In the last line of the poem, the word “sound” is printed “ <i>sori</i> 소리” (pg. 57).
2	120	No <i>mo chŏm</i> appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven.	No <i>mo chŏm</i> appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven.	A smudge/ <i>mo chŏm</i> appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven.	A smudge/ <i>mo chŏm</i> appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven.	A comma appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven (pg. 198).	A comma appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven (pg. 344).	A comma appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven (pg. 113).	A comma appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven (pg. 117).	A comma appears between <i>pyŏgae</i> 비개 and <i>hŭryŏnan</i> 가 흐릿난가 in line seven (pg. 80).
3	126	In the first line on the page “ <i>talpit</i> 달빛” appears normally.	In the first line on the page “ <i>talpit</i> 달빛” appears normally.	In the first line on the page “ <i>talpit</i> 달빛” appears as <i>malpit</i> 말빛.” This appears to be an artifact of the process of photocopying.	In the first line on the page “ <i>talpit</i> 달빛” appears as <i>malpit</i> 말빛.” This appears to be an artifact of the process of photocopying.	“ <i>Talpit</i> 달빛” appears as “ <i>malpit</i> 말빛.” In a footnote, Kwŏn writes “Kim Yong-jik suggests this is an error” (pg. 205).	“ <i>Talpit</i> 달빛” appears as “ <i>malpit</i> 말빛” (pg. 354). Kim has a footnote that says <i>mal</i> is an error in the “original” (pg. 367).	“ <i>Talpit</i> 달빛” appears as “ <i>malpit</i> 말빛” (pg. 117). Kim has a footnote that says <i>mal</i> is an error (pg. 177).	“ <i>Talpit</i> 달빛” appears as “ <i>malpit</i> 말빛” (pg. 119). Kim has a footnote that says <i>mal</i> is an error (pg. 435).	“ <i>Talpit</i> 달빛” appears as “ <i>talbit</i> 달빛” (pg. 81).



Bibliography

“List Makers”

Enumerative

Descriptive



Information Science

Patrick Wilson *Two Kinds of Power: An Essay on Bibliographical Control* (1968)

C. 1-4
Law
Office
(S.C.-5)

South Carolina (Colony) Court of Chancery.
Records of the Court of Chancery of South Carolina,
1671-1779; edited by Anne King Gregorie, with an introd.
by J. Nelson Frierson. Washington, American Historical
Association, 1950.
676 p. 26 cm. (American legal records, v. 6)

1. South Carolina—Hist.—Colonial period—Sources. B. Chancery.
South Carolina

Leader 01034cam a2200277a 4500
001 99002459390203941
005 2016L1109021656.0
008 7602271950 dcu b 001 0 eng c
010 ##\$aa**52009246*
035 ##\$a[OCoLC]20176275[OCoLC]221260455[OCoLC]5605193455[OCoLC]7713643185[OCoLC]9764668545[OCoLC]1001185451
040 ##\$bDetroit. Public Library\$CTC\$DLSU\$DHL5
043 ##\$an-us:c
050 #45af2725b.567
090 ##\$af2725b.567
110 1#5aSouth Carolina.SbCourt of
Chancery\$Dhttp://id.loc.gov/authorities/names/n8807180750http://viaf.org/viaf/sourceID/LC/n8807180752LCNAME5041-LIBRARY_OF_CONGRESS-n
8807180759Y
245 10\$aRecords of the Court of Chancery of South Carolina, 1671-1779 /\$cedited by Anne King Gregorie ; with an introduction by J. Nelson Frierson.
260 ##\$aWashington, D.C. :\$bAmerican Historical Association,\$c1950.
300 ##\$a676 p. :\$c26 cm.
490 1\$aAmerican legal records ;\$v. 6
504 ##\$aincludes bibliographical references and index.
648 #75a1600-17750http://id.worldcat.org/fast/fst0135564152FAST5041-OCLC-fst0135564159Y
648 #75aColonial Period (United States)\$2FAST5041-OCLC-fst0135564159N
650 #05aCourts\$South Carolina\$History\$0http://id.loc.gov/authorities/subjects/sh8503357152LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159Y
650 10\$aEquity pleading and procedure\$2South Carolina\$Cases\$0http://id.loc.gov/authorities/subjects/sh8504455952LCSH5041-
LIBRARY_OF_CONGRESS-sh 8504455959Y
650 #05aLaw reports, digests, etc.\$2South Carolina\$0http://id.loc.gov/authorities/subjects/sh200912929652LCSH5041-LIBRARY_OF_CONGRESS-
sh200912929659Y
650 #05aCourt records\$2South Carolina\$0http://id.loc.gov/authorities/subjects/sh850335652LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599Y
650 #05aCourts\$Records and correspondence\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599N
650 #05aRecords of courts\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599N
650 #05wg\$Archives\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599R
650 #05wg\$Evidence (Law)\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599R
650 #05wg\$Public records\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503356599R
650 10\$aChancery\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959N
650 10\$aEquity pleading and procedure\$2Law and legislation\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959N
650 10\$wg\$Civil procedure\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959R
650 10\$wg\$Equity\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959R
650 10\$wg\$Pleading\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959R
650 10\$wg\$Procedure (Law)\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959R
650 10\$wg\$Trial practice\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8504455959R
650 #05aCourts\$Law and legislation\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159N
650 #05aJudiciary\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159N
650 #05wg\$Dispute resolution (Law)\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05wg\$Judicial districts\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05wg\$Law\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05wg\$Procedure (Law)\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05aJudicial power\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05aJurisdiction\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
650 #05aJustice, Administration of\$2LCSH5041-LIBRARY_OF_CONGRESS-sh 8503357159R
651 #05aSouth Carolina\$History\$Colonial period, ca. 1600-
1775\$Sources\$0http://id.loc.gov/authorities/names/n7902291450http://viaf.org/viaf/sourceID/LC/n7902291452LCNAME5041-LIBRARY_OF_CONGRESS-
n 7902291459Y
651 #05aSouth Carolina (Colony)\$2LCNAME5041-LIBRARY_OF_CONGRESS-n 7902291459N
651 #05aSouth Carolina (Province)\$2LCNAME5041-LIBRARY_OF_CONGRESS-n 7902291459N
651 #05aIshnaia\$Karolina\$2LCNAME5041-LIBRARY_OF_CONGRESS-n 7902291459N
655 #75aAbstracts\$0[OCoLC]fst0142368350http://id.worldcat.org/fast/fst0142368352FAST5041-OCLC-fst0142368359Y
655 #75aHistory\$0[OCoLC]fst0141162850http://id.worldcat.org/fast/fst0141162852FAST5041-OCLC-fst0141162859Y
655 #75aSources\$0[OCoLC]fst0142390050http://id.worldcat.org/fast/fst0142390052FAST5041-OCLC-fst0142390059Y
655 #75aTrials, litigation, etc.\$0[OCoLC]fst0142371250http://id.worldcat.org/fast/fst0142371252FAST5041-OCLC-fst0142371259Y
655 #75aCompendia\$2FAST5041-OCLC-fst0142368359N
655 #75aCompendiums\$2FAST5041-OCLC-fst0142368359N
655 #75aCompendis\$2FAST5041-OCLC-fst0142368359N
655 #75aEpitomes\$2FAST5041-OCLC-fst0142368359N
655 #75aPreliis\$2FAST5041-OCLC-fst0142368359N
655 #75aPublication summaries\$2FAST5041-OCLC-fst0142368359N
655 #75aSynopses (Abstracts)\$2FAST5041-OCLC-fst0142368359N
655 #75aSynopses (Abstracts)\$2FAST5041-OCLC-fst0142368359N
655 #75wg\$Derivative works\$0[OCoLC]fst0191991752FAST5041-OCLC-fst0142368359R
655 #75wg\$Informational works\$0[OCoLC]fst0191993052FAST5041-OCLC-fst0142368359R
700 1#5aGregorie, Anne King\$0http://id.loc.gov/authorities/names/n8807180650http://viaf.org/viaf/sourceID/LC/n8807180652LCNAME5041-
LIBRARY_OF_CONGRESS-n 8807180659Y
710 2#5aAmerican Historical
Association\$0http://id.loc.gov/authorities/names/n7906358350http://viaf.org/viaf/sourceID/LC/n7906358352LCNAME5041-LIBRARY_OF_CONGRESS-n
7906358359Y
710 2#5aHA\$2LCNAME5041-LIBRARY_OF_CONGRESS-n 7906358359N
776 08\$Online version:\$aSouth Carolina. Court of Chancery.\$tRecords of the Court of Chancery of South Carolina, 1671-1779.\$dWashington, D.C. :
American Historical Association, 1950.\$w[OCoLC]758496938
830 #05aAmerican legal records\$svv. 6\$0http://id.loc.gov/authorities/names/n8370570050http://viaf.org/viaf/sourceID/LC/n8370570052LCNAME5041-
LIBRARY_OF_CONGRESS-n 8370570059Y
900 ##\$a002459399
901 7#5aAMC2250
903 ##\$aMHAMC22507HU
905 ##\$af**0272
948 ##\$bORC170ASWID2boc032535b199603215c19960322
991 ##\$b2022-01-11\$cWorldCat record variable field(s) change: 700
991 ##\$b2022-06-12\$cWorldCat record encoding level change - WorldCat record variable field(s) change: 655
998 ##\$aNO-OVERLAY
AVA ##\$099002459399020394182218517867900039415a01HVD_INST5bWIDScOld Widener\$duS 19596.715eavailable\$f15g05GENSKZHC1Sp4Widener
Library
AVA ##\$099002459399020394182218517867700039415a01HVD_INST5bLAWScOffsite Storage\$dkfS1849 .C5x
1850\$available\$f15g05jHDSK05p2SqHarvard Law School Library
CAT ##\$aHistory_Archaeology\$bRegions_Countries\$cAmericas_United_States_Local_History
INST ##\$a01HVD_INST
INT ##\$af

Compare BIBFRAME converted to MARC

Identifier Bib ID LCCN
 Serialization XML Text

BIBFRAME XML

```

<rdf:RDF xmlns:bf = "http://id.loc.gov/ontologies/bibframe/" xmlns:bfcl = "http://id.loc.gov/ontologies/bfcl/" xmlns:madsrdf =
"http://www.loc.gov/mads/rdf/v1#" xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs = "http://www.w3.org/2000/01/rdf-
schema#" >
<bf:Instance rdf:about="http://id.loc.gov/resources/instances/20987659" >
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Print" />
<bf:issuance >
<bf:Issuance rdf:about="http://id.loc.gov/vocabulary/issuance/mono" >
<rdfs:label >single unit</rdfs:label>
</bf:Issuance>
</bf:issuance>
<bf:provisionActivity >
<bf:ProvisionActivity >
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />
<bf:date rdf:datatype="http://id.loc.gov/datatypes/edtf" >1950</bf:date>
<bf:place >
<bf:Place rdf:about="http://id.loc.gov/vocabulary/countries/dcu" >
<rdfs:label xml:lang="en" >District of Columbia</rdfs:label>
</bf:Place>
</bf:place>
</bf:ProvisionActivity>
</bf:provisionActivity>
<rdfs:label >Records of the court of chancery of South Carolina 1671-1779 :</rdfs:label>
<bf:responsibilityStatement >Edited by Anne King Gregorie</bf:responsibilityStatement>
<bf:title >
<bf>Title >
    
```

MARC (Text)

```

cam a22 5c 4500
2 $aSouth Carolina$http://id.loc.gov/rwo/agents/n79022914$http://id.loc.gov/rwo/agents/n79022914$http://id.lo
c.gov/vocabulary/relators/ctb
001 20987659
003 DLC
005 20190530144430.0
008 190528s1950||||dcu 00| 0|eng
040 $aDLC$beng$erda$cDLC
084 $aLAW U.S. South Carolina 5 1950$qdlc
245 10 $aRecords of the court of chancery of South Carolina 1671-1779$cEdited by Anne King Gregorie
264 1 $aWashington, D.C.$bThe American Historical Association$c1950
300 $a6, 676 pages$c26 cm
334 $asingle unit$bmono$0http://id.loc.gov/vocabulary/issuance/mono
336 $atext
337 $aunmediated
338 $avolume
490 0 $aAmerican Legal Records ; Vol. 6
884 $aDLC bibframe2marc v2.3.0 (MarkLogic Corporation)$g20231028205518.0$qDLC$uhttps://github.com/lcnetdev/bibf
rame2marc
    
```

Conversion: [DLC bibframe2marc v2.3.0](#)

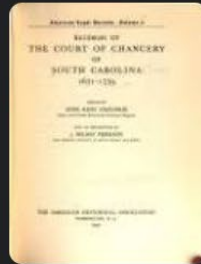
Records of the Court of Chancery of South Carolina, 1671-1779

By South Carolina. Court of Chancery · 1950

Preview unavailable

Search inside

+ Add to my library



Preview unavailable

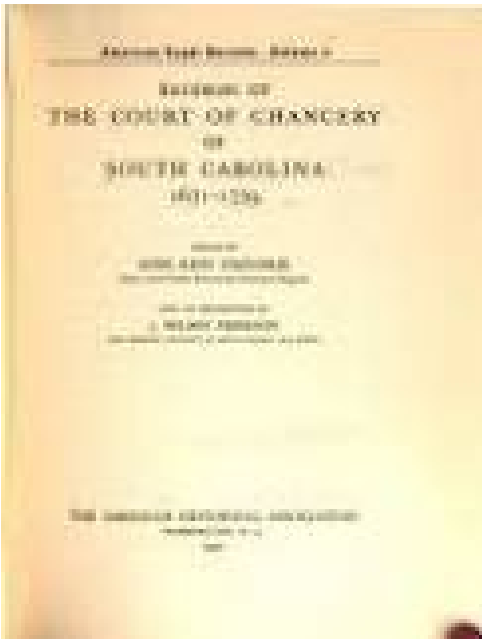
[Learn more](#)

Overview

Get the book

Publisher collection

Google Books



THE UNIVERSITY OF IOWA LIBRARIES
INFOHAWK+



E-BOOK

Records of the Court of Chancery of South Carolina, 1671-1779 / edited by Anne King Gregorie ; with an introduction by J. Nelson Frierson.

South Carolina. Court of Chancery, court governed. ; Anne King Gregorie editor.; American Historical Association, issuing body.
Washington, D.C. : American Historical Association 1950

[Available at Law Library Stacks \(KFS2315 .A545 1671\) >](#)

[Online access may be available >](#)



複
寫

複 복 겹칠 (overlap, compound)
寫 사 베끼다 (copy, imitate)

베
끼
다

COPY

- I. A. ABUNDANCE, PLENTY.
B. FULLNESS, PLENTITUDE.
- II. A TRANSCRIPT OR REPRODUCTION
OF AN ORIGINAL.

Copy serving
information
retrieval

Distance of some kind
(e.g. **space**)

Copy

Distance of some kind
(e.g. *space*)

Copy

Copy

Distance of some kind
(e.g. **time**)

Transform. 化.

Latin *transformāre*, < [TRANS- prefix](#) + *formāre* to form>
trans, 'across, to or on the farther side of, beyond, over'
化 to become, to change

Accounting

I. To render a reckoning.



Etymology: < Anglo-Norman *acounter*, *aconpter*, *accomptier*, *accomptere*, Anglo-Norman and Old French *acunter*, Anglo-Norman and Old French, Middle French *aconter*, Middle French *acompter*, *acompter* **to count, enumerate (c1100), to relate, tell, to include (something) in a reckoning, count in, to calculate (all early 12th cent.), to consider, deem, to value, consider important (c1280), to present financial accounts (early 14th cent.), to matter, be of account** (mid 14th cent.) < a- [A- prefix](#) + *cunter*, *conter* [COUNT v.](#) Compare Old Occitan *acomtar*, *acontar* (c1150), post-classical Latin *acomputare* (1661 in a British source). The [β. forms](#) reflect Anglo-Norman and Middle French *acompter*, *acompter*, etc.: for the development of forms of this type and their history in French see discussion at [COUNT v.](#)

Copy

Distance of some kind (e.g. time and/or space)

Copy

Similarity,
similitude

Particularity,
Individuality

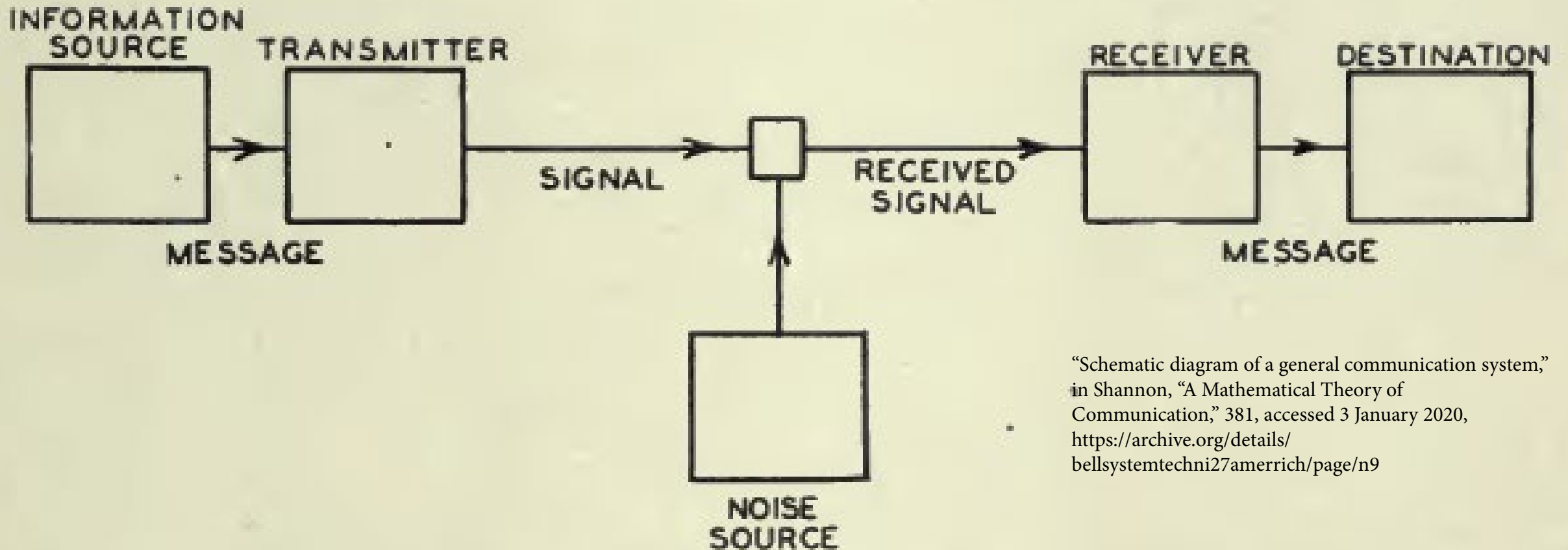
Who or what determines?

For what purposes?

In which contexts?

COPIES AND CLAUDE SHANNON

Shannon's theory of communication became known as "information theory." As a practical matter it is about making a copy at a distance. It would better be named Copy Theory . . .



"Schematic diagram of a general communication system," in Shannon, "A Mathematical Theory of Communication," 381, accessed 3 January 2020, <https://archive.org/details/bellssystemtechni27amerrich/page/n9>

Deep Learning

As Bibliographical Processes

Input space

y

“Gradient”
of steep

Which direction
 $C(x, y)$ most d

-3 -2 -1

1

(x, y)

$C(w)$

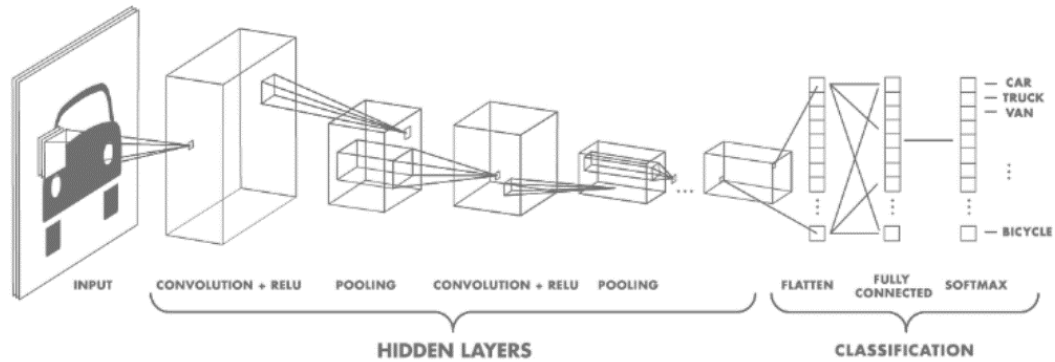
copies and
accounting for
them facilitates

DEEP LEARNING

AUTOMATED PATTERN RECOGNITION

w

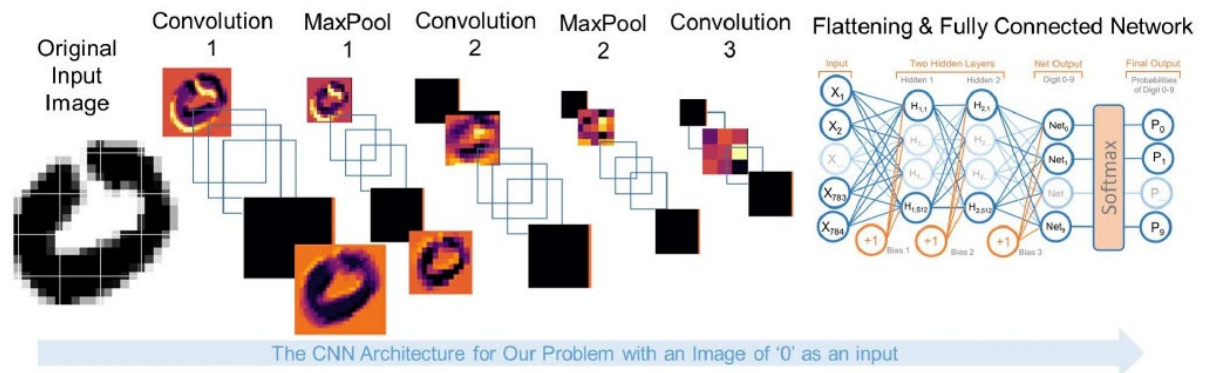




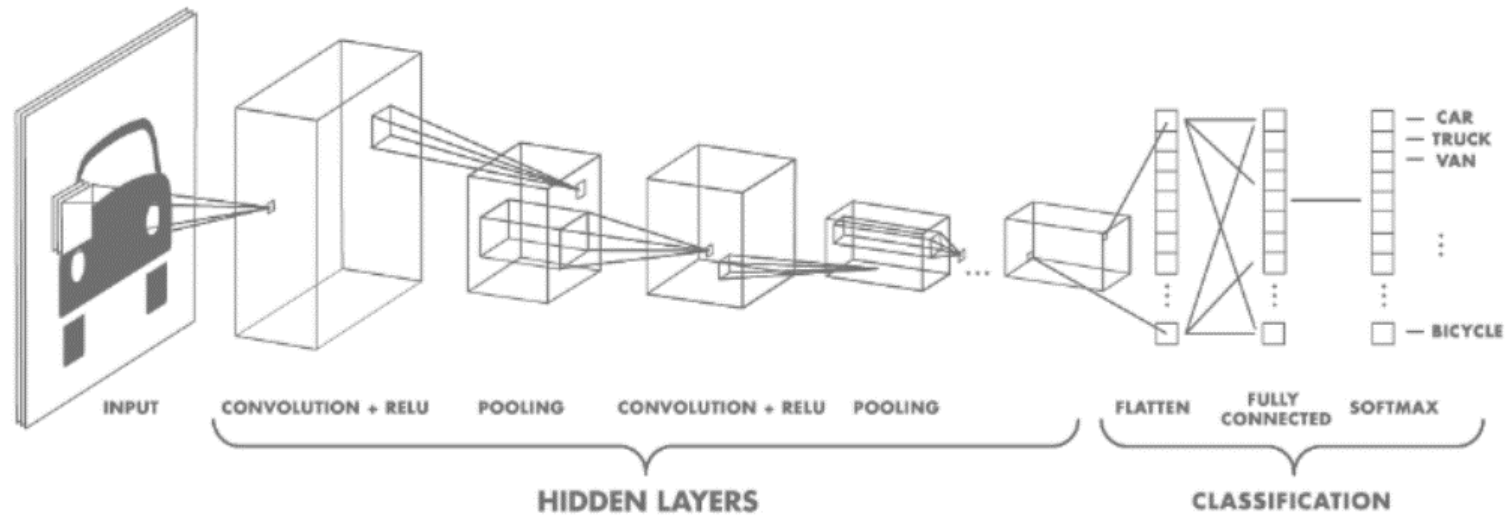
Deep Learning

Architecture of a CNN.—Source: <https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html>

Convolutional Neural Networks

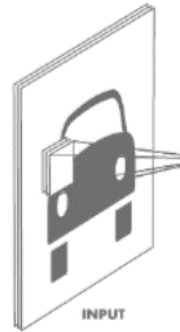


Convolutional Neural Networks



Architecture of a CNN.— Source: <https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html>

Convolution



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

Input

plus

1	0	1
0	1	0
1	0	1

Filter / Kernel

1x1	1x0	1x1	0	0
0x0	1x1	1x0	1	0
0x1	0x0	1x1	1	1
0	0	1	1	0
0	1	1	0	0

4		

1	1x1	1x0	0x1	0
0	1x0	1x1	1x0	0
0	0x1	1x0	1x1	1
0	0	1	1	0
0	1	1	0	0

4	3	

1	1	1x1	0x0	0x1
0	1	1x0	1x1	0x0
0	0	1x1	1x0	1x1
0	0	1	1	0
0	1	1	0	0

4	3	4

1	1	1	0	0
0x1	1x0	1x1	1	0
0x0	0x1	1x0	1	1
0x1	0x0	1x1	1	0
0	1	1	0	0

4	3	4
2		

1	1	1	0	0
0	1x1	1x0	1x1	0
0	0x0	1x1	1x0	1
0	0x1	1x0	1x1	0
0	1	1	0	0

4	3	4
2	4	

1	1	1	0	0
0	1	1x1	1x0	0x1
0	0	1x1	1x0	1x0
0	0	1x1	1x0	0x1
0	1	1	0	0

4	3	4
2	4	3

1	1	1	0	0
0	1	1	1	0
0x1	0x0	1x1	1	1
0x0	0x1	1x0	1	0
0x1	1x0	1x1	0	0

4	3	4
2	4	3
2		

1	1	1	0	0
0	1	1	1	0
0	0x1	1x0	1x1	1
0	0x0	1x1	1x0	0
0	1x1	1x0	0x1	0

4	3	4
2	4	3
2	3	

1	1	1	0	0
0	1	1	1	0
0	0	1x1	1x0	1x1
0	0	1x0	1x1	0x0
0	1	1x1	0x0	0x1

4	3	4
2	4	3
2	3	4



Useful
similarity

4	3	4
2	4	3
2	3	4

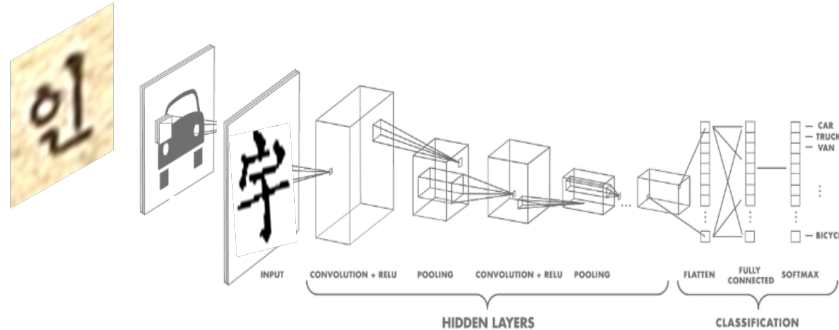
Feature
(map)

a
copy

Deep Learning

copies (like turtles) all the way down

가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가	가
각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각
각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각	각
각	각	각	각	각	각	각	간	간	간	간	간	간	간	간	간	간	간



Weight File

(record of which features best predict/ describe objects in training set)

Something.weights

car

99 87 114

236 157 184



229 173 151



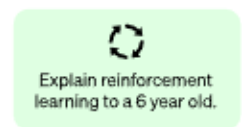
Enumeration Description

Analysis Critique

Step 1

Collect demonstration data and train a supervised policy.

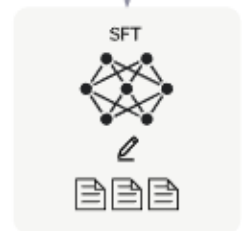
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



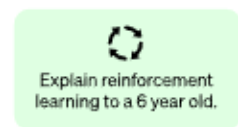
This data is used to fine-tune GPT-3.5 with supervised learning.



Step 2

Collect comparison data and train a reward model.

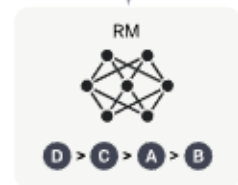
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

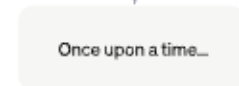
A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.



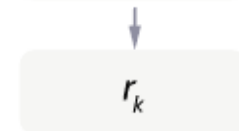
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



opportunities for deep learning

Early-to-Mid Twentieth-Century Korean Periodicals

野 談
創刊號



十二月號



學 生 界

第八號



行發 社會式株書圖城漢 城京

中 學 講 義 錄

第二十號

朝鮮通信中學館

納本朝
朝鮮總督府警務局
大正
年 月 日
號

大正十年十一月十九日約東郵便物認可

檢閱者

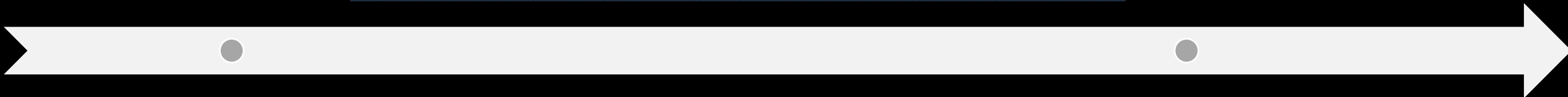
89,981 images

1,009 rare
(*kwijungbon*)
periodical issues

932 colophons
identified & transcribed
automatically

1910

1960



一九四六年二月廿日印刷 第一號
一九四六年三月一日發行

(本號定價拾圓)

(送料二〇錢)

編輯兼 發行人 金 東 鎮
서울市御成町八六 凡閣章內

印刷所 朝鮮精版社
서울市長谷川町七四

定價普通號 一部 拾圓 送料二〇錢
半年分六十圓 (但、先金・送料添付)
增大號日時七超過額을別受함

駐文規定

★本誌의 註文은 一切先金으로 하시오
(送料添付)

★送金은 반드시 小爲管를 利用 하시오

(서울市御成町八六番地에 支拂)

★一切의 問議는 郵票同封或은 往復葉書
로 하시오

★郵票代用은 一切 拒絕합니다

發行所 서울市御成町 凡章閣
八十六番地

電話本局②四六五九

물음 (郵票添付要)

○本誌를 讀고 지하시든 本은 地方支局에
미리 注文豫約하여 주십시오 部數의
制限이 있을므로 늦으면 希望者全
部에게 드리기 困難합니다.

○萬一 地方에 支局이 없어서 不便한
處는 直接本社에 先金으로 豫約하
여 주십시오. 送料 二圓을 添付하
여 振替 京城二〇二又是 小爲替切
手代用 予었어면지 樂합니다.

西紀一九四七年六月二十五日印刷
西紀一九四七年七月一日發行

서울市乙支路二街一八〇番地
編輯兼 發行人 金正修

印刷所 學生社印刷部

서울市乙支路二街一八〇番地
發行所 學生社

編輯局 電(本)六一六番
電話本局九一九番
振替 京城二〇二番
臨時定價五拾圓也

(購讀料先金)
一個月 十錢
六個月 五十錢
一年 一圓〇五錢

廣告料
十七字一行 三十錢
半頁 一圓八十錢
一頁 三圓
一頁以上은 (每頁)
一圓五十錢式加算

大正五年二月二十五日
大正五年三月五日

印刷 發行

京城府敦義洞百三十一番地

丁元燮

著作兼 發行所 京城府勸農洞三十一番地

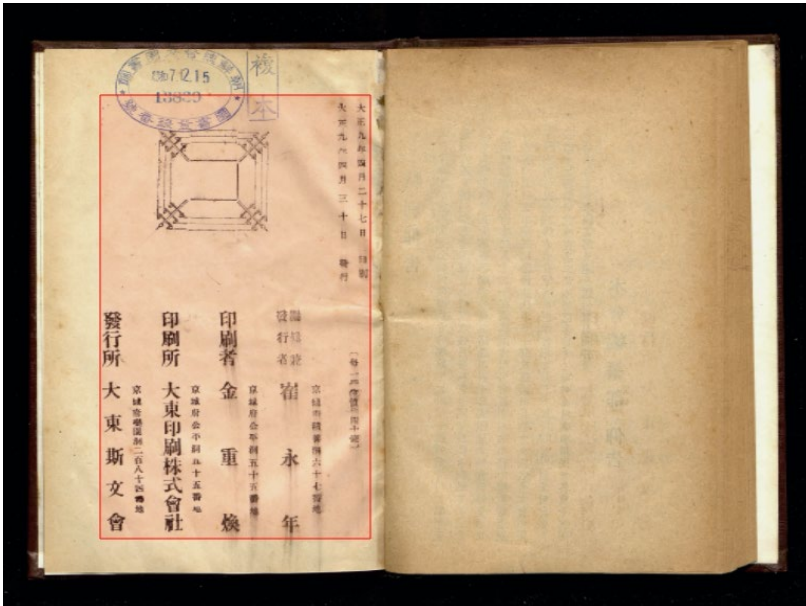
印刷者 韓養浩

京城府鍾路通一丁目三九番地

印刷所 鮮明社

京城府嘉會洞百七十七番地

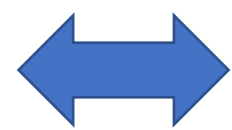
發行所 侍天教報社



Section Model

4	3	4
2	4	3
2	3	4

Features



target

colophon

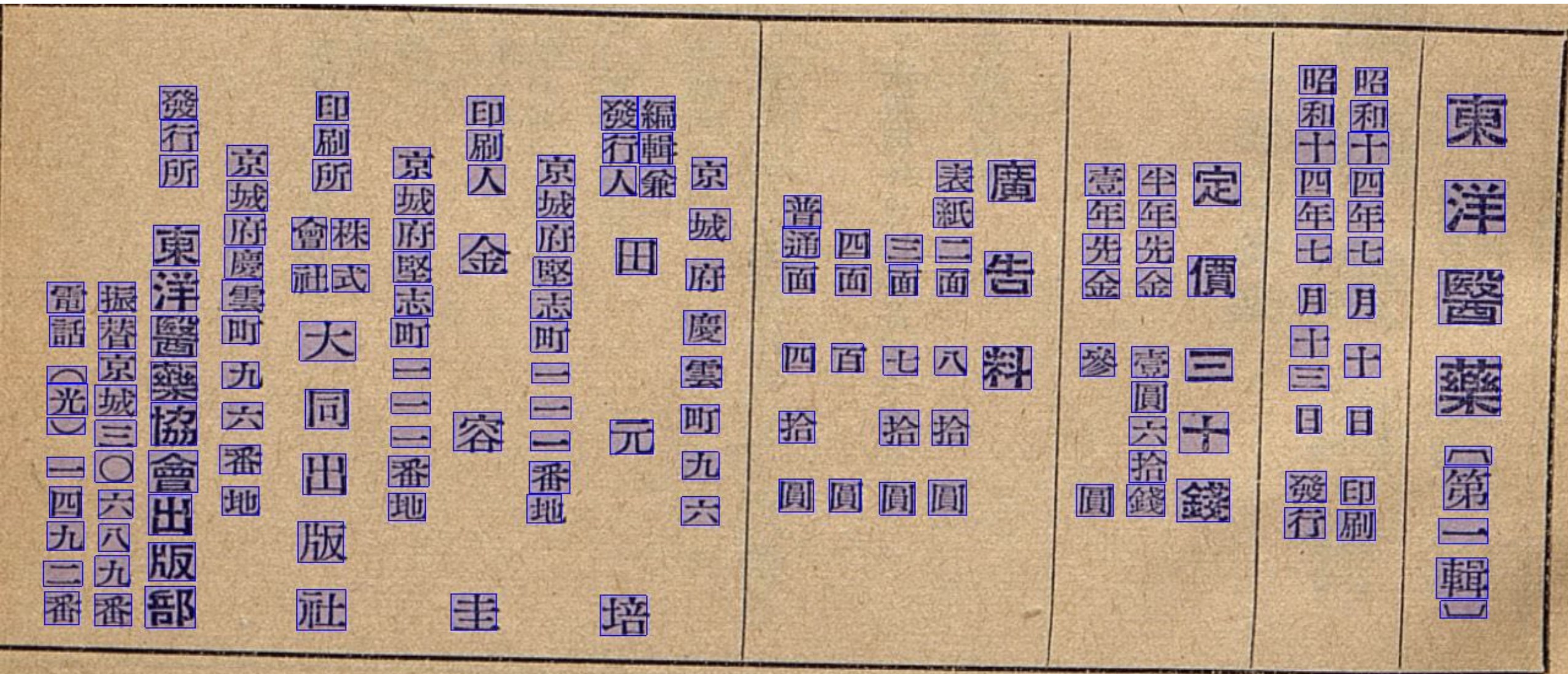
4	3	4
2	4	3
2	3	4

Features



target
meaningful
element

Segmentation Model



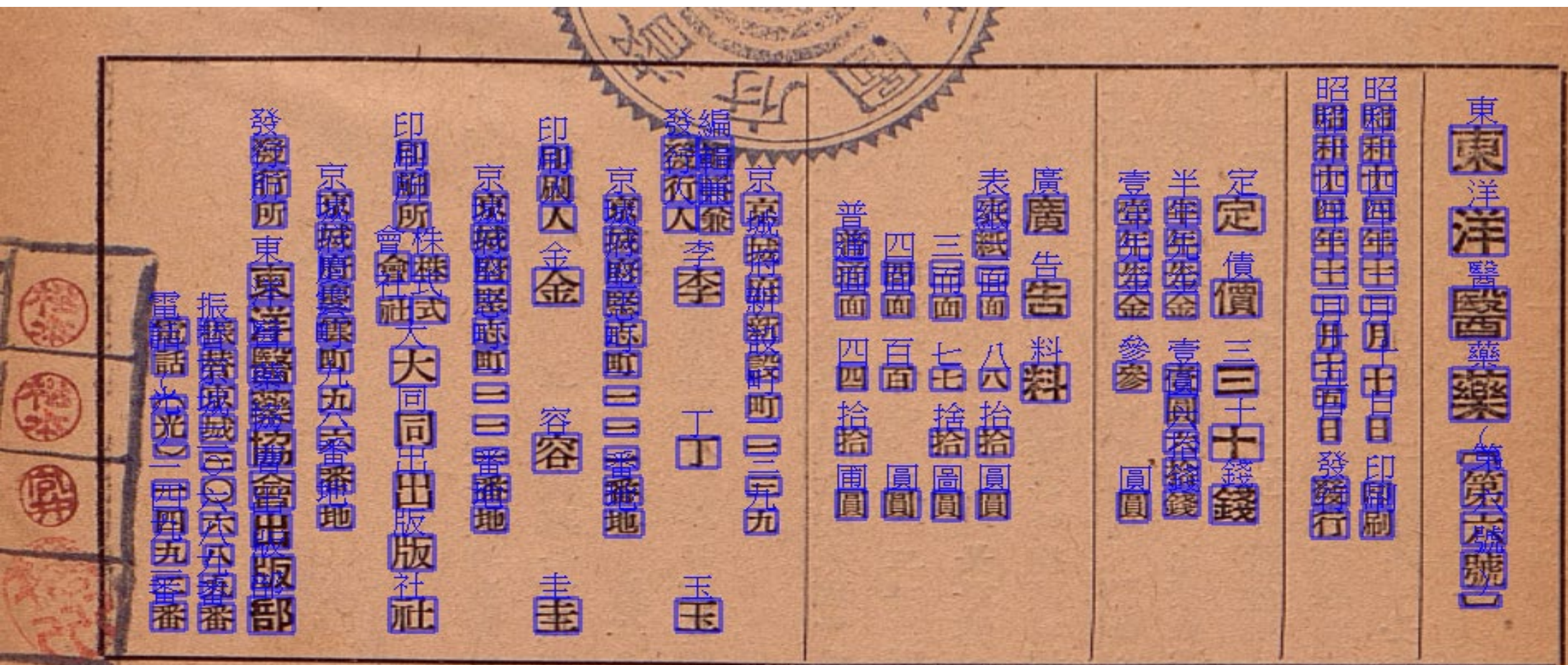
4	3	4
2	4	3
2	3	4

Features

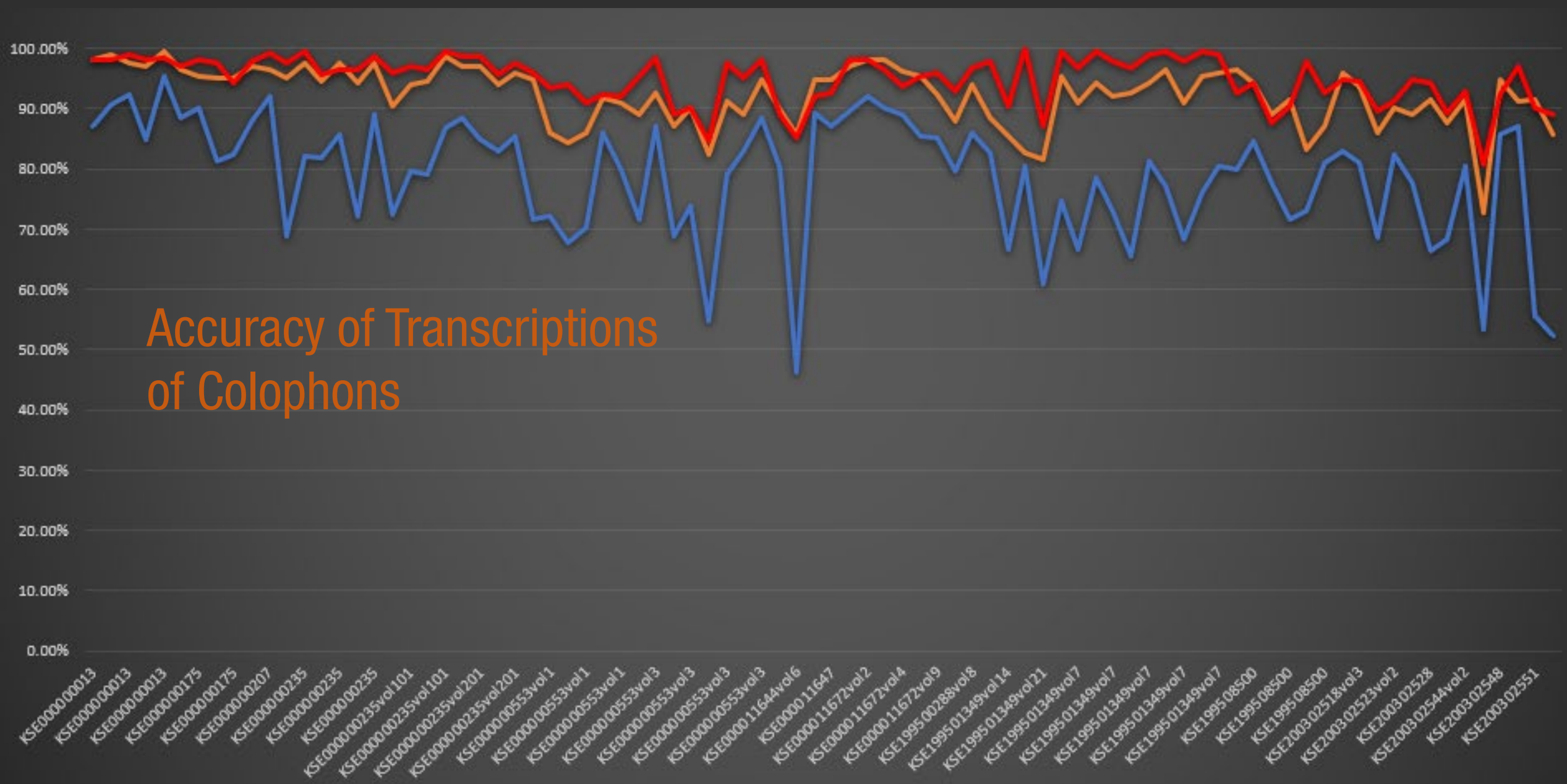


target
labels for
meaningful
elements

Classification Model

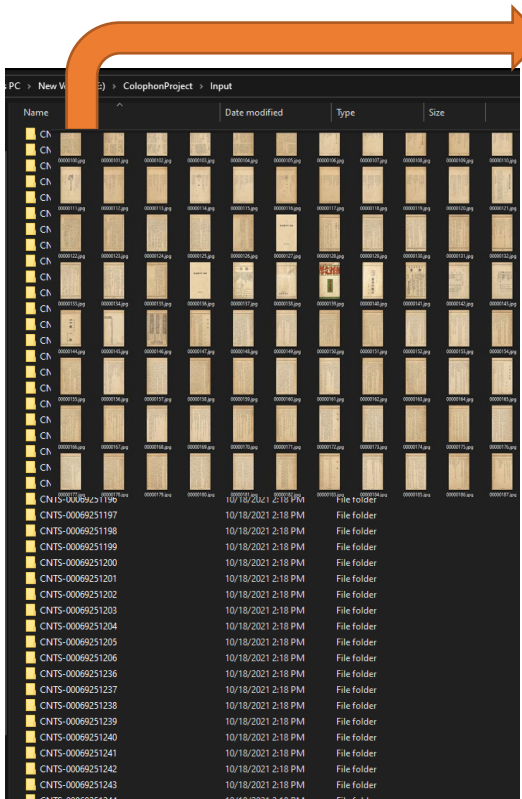


Model 2 (Avg. accuracy 78.6%) Model 3 (Avg. accuracy 92.7%) Model 4 (Avg. accuracy 95%)



Accuracy of Transcriptions of Colophons

Section Model Segmentation Model Classification Model



MooN Process Manager

INPUT: E:\WColophonProject\Winput

OUTPUT: E:\WColophonProject\WOutput

Section: section_seg_01

Segmentation: OldNewSeg5_Grid7

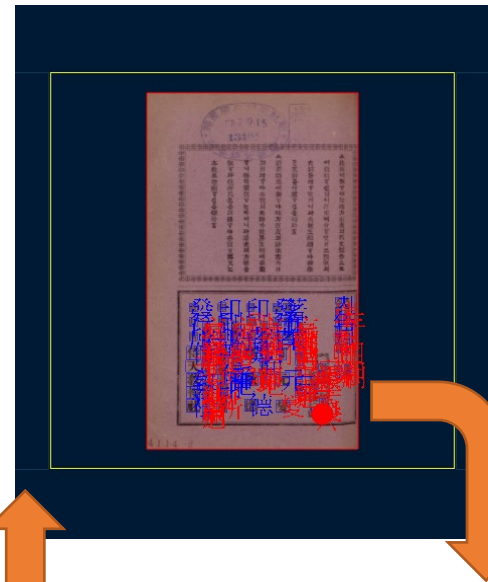
Classification: oldchar_9895

프로세스 추가

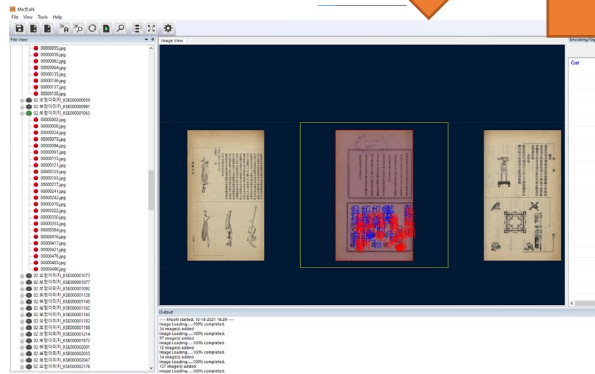
작업 리스트: 548개, 인식중: 0개

No.	시작	정지	인식 프로세스 생성 전
No.1	시작	정지	인식 프로세스 생성 전
No.2	시작	정지	인식 프로세스 생성 전
No.3	시작	정지	인식 프로세스 생성 전
No.4	시작	정지	인식 프로세스 생성 전
No.5	시작	정지	인식 프로세스 생성 전
No.6	시작	정지	인식 프로세스 생성 전
No.7	시작	정지	인식 프로세스 생성 전
No.8	시작	정지	인식 프로세스 생성 전

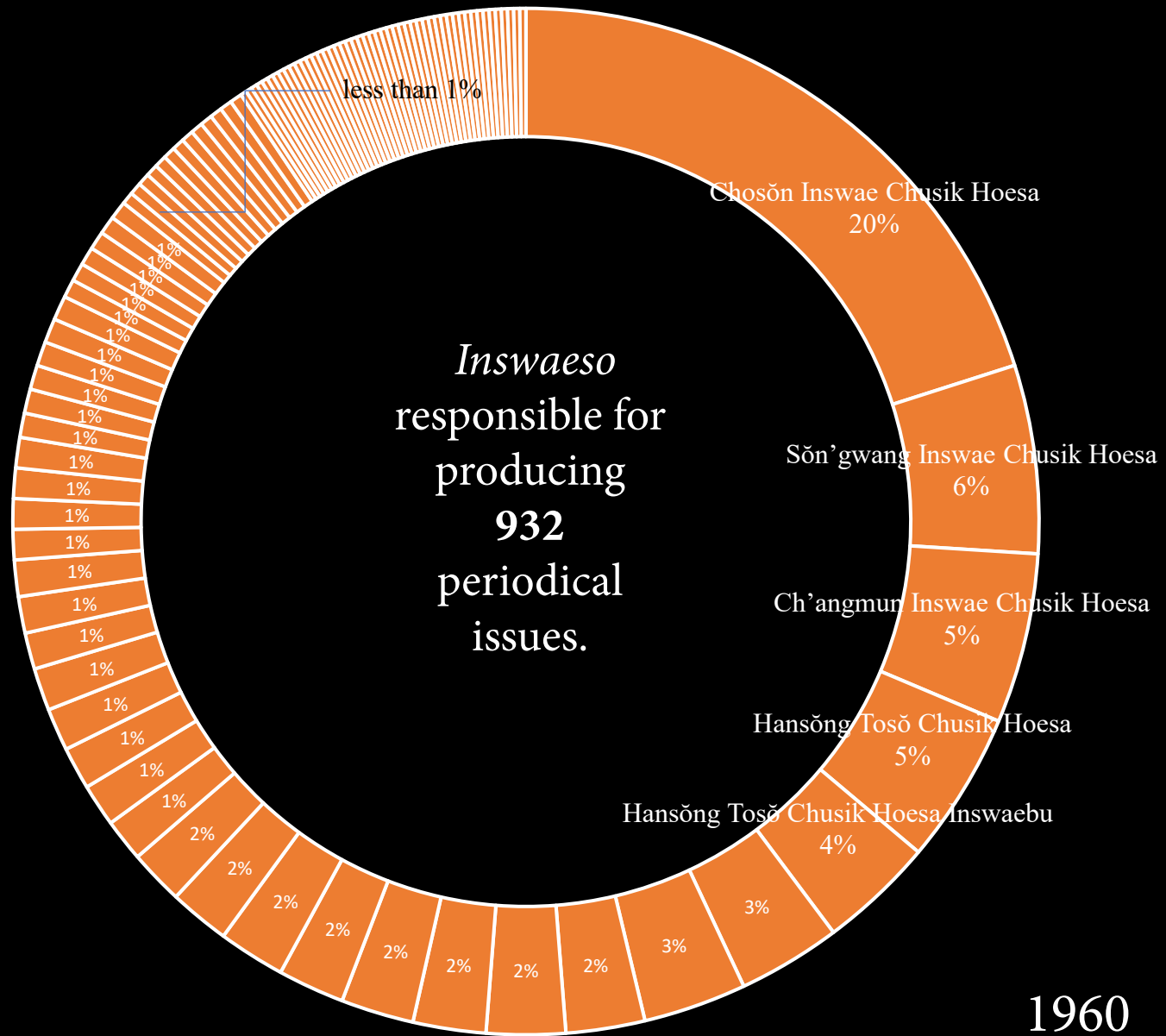
완료



00000242.Jpg



- 1 年 :+相 印刷
- 2 치 遼年二月劃卿 發가
- 3 定價傘 五錢穴
- 4 獎費 ●
- 5 京兼脚鄰會廻百平三番地
- 6 著作兼
- 7 發行者 丁元變
- 8 京坡府薊廻三面 爾地
- 9 印刷者 鄭敬德
- 10 京所府苑痾三暝番地
- 11 印刷所 朝鮮福瘡印한所
- 12 京撼府嘉覽載帝g七五地奮廉廻
- 13 發行所 侍天教報社



1910

1960

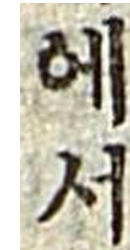
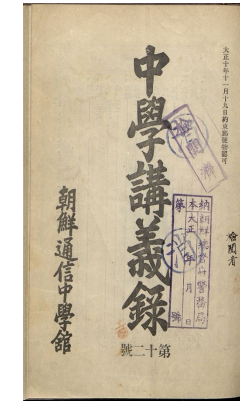
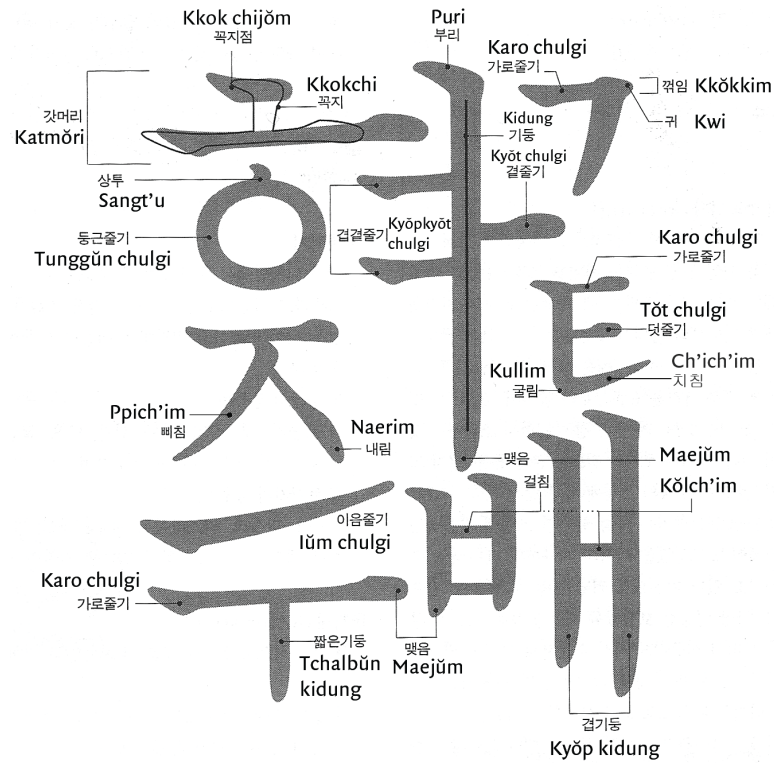


Newly built
Hansŏng Tosŏ
Chusik Hoesa
building,
Haksaenggye
(December
1920),
unnumbered
front matter
(image from
microfilm at
the NLK).

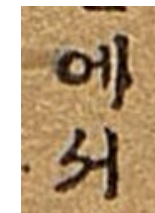
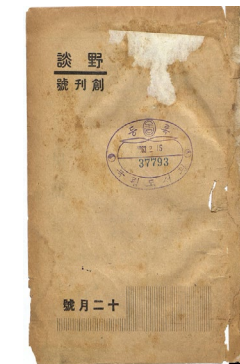


Type samples

from materials printed by Hansöng Tosö Chusik Hoesa



1920s

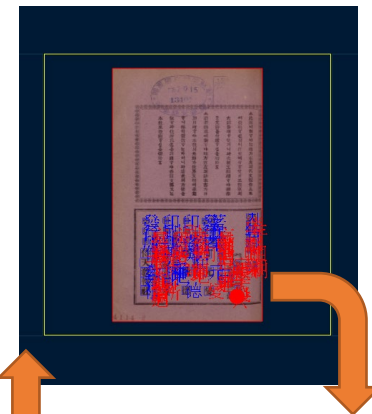
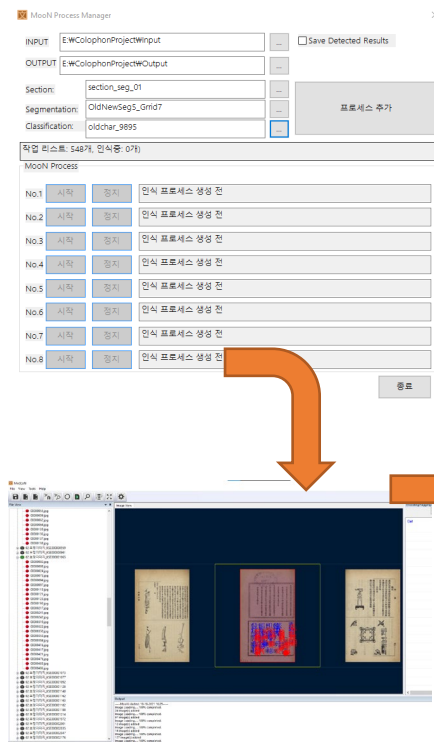
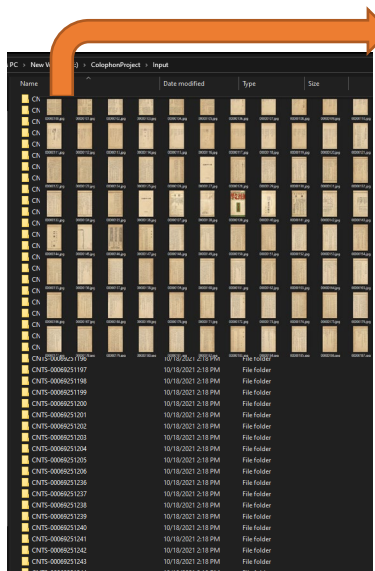
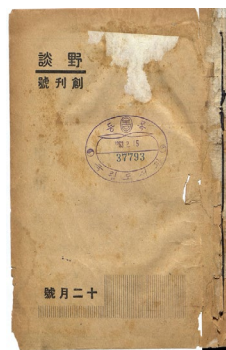
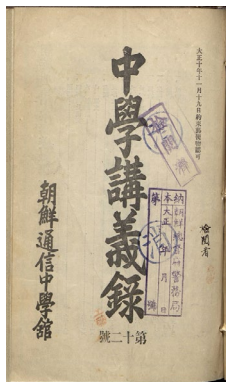
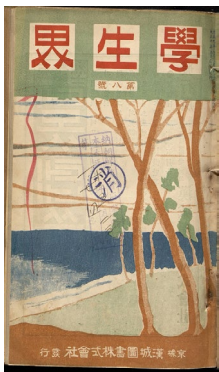


1930s

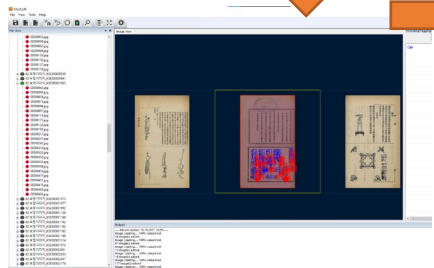
Source: Han Chae-jun, ed., *Han'gül küllkol yongö sajön* (A dictionary of typographic terms for han'gül) (Seoul: Sejong Taewang Kinyöm Saöphoe, 2000), 46.

Type samples from page 101 of the May 1921 issue of *Haksaenggye* (top left), page 78 of the August 1922 issue of *Chunghak Kangürok* (top right), page 72 of the June 1935 issue of *Ch'örhak* (bottom left), and page 157 of the November 1937 issue of *Yadam* (bottom right).

Full text transcriptions of the Periodicals held by the National Library of Korea



00000242.Jpg



세계를 한 집으로 보고 사해를 한 형제로 본다 하면 말이
갓고 풍속이 갓고 역사가 갓고 이해가 가튼 한 민족은
더욱이나 한 집안 한 형제로 보게 될 것입니다. 설사 하는
일은 세계 전체를 표준한다 할지라도 세계가 하루 아침에
한 집안 한 형제가 되어 가지고 세계 전체의 일을 할 수가
업고 민족과 민족이 어울려져 내 일도 전체의 일이오 네
일도 전체의 일이라 하는데서 其實 사해 형제의 본의가
실현될 줄로 생각합니다. 그것은 오랫동안 관계를 가져온
한 민족이 스스로 어울려지기 쉽고 관계맺기 쉬울 경우에
잇는 까닭이 안이갓습니까. 이 점에서 한 가지로 우리 朝鮮
동포라 하면 그 하는 일의 표준을 세계적이거나
민족적이거나를 물론하고 스스로 동포와 동포가 손목을
잡고 일터에 나아가게 되는 것입니다. 더욱이 그 민족의
일을 그 민족이 자진하여 하지 안이하면 즉 자기의 맞은
일을...

more opportunities for deep learning

Qisha Canon
磧砂藏

Qisha Canon Encoding Project (Fo Guang Temple, Taiwan)

遠伏八勤勤掛尼弘恭我軀氏澤玄於兜衛稱邪離
 一伺六北品報孝孤山影悟截斯新求妙珍立罪花西誦辦重
 七但其十截塗孫岸影戲斷漆汝為珠現竟義芻獲
 三位具千報孤山影戲斷漆汝為珠現竟義芻獲
 上住冥南唐塵學芥佛悲戲方覺沈
 下何決印唯境穴樂得惡承
 不佗佛印唯境穴樂得惡承
 出世作佛印唯境穴樂得惡承
 並使了及喧壤之已從想捨是樂煩瑣春粉耶解觀退門響
 乃供切支喻士宣布復愁捨標橋洲眼後簡耳莊笑解觀退門響
 久依初取僕外家帝循微意提智燭生著好聚華言記變遠進
 之便初取僕外家帝循微意提智燭生著好聚華言記變遠進
 乘俗判受器多察心愛揭暴欠海燈四鑄素豎聽隆訶貧
 九信利句嚴密常忍慈提暴欠海燈四鑄素豎聽隆訶貧
 也修到各因天寶識志暮損晚欲淨燭男礙終脫能業誦貧
 亂俱則同固天寶識志暮損晚欲淨燭男礙終脫能業誦貧
 了滿前各向夫失專幹念慧推日歡歡深燭男礙終脫能業誦貧
 事假別向獨夫失專幹念慧推日歡歡深燭男礙終脫能業誦貧
 二獨剛君國專契底念愛擇曾正清者異神維脫能業誦貧
 云健劇吠園團對度性憐攝最歸滅渴犯疑當禁網緊自餐誦貧
 五傾功告圓奪導小摩待意放有死湖狀疾乘稟總自餐誦貧
 亦傾功告圓奪導小摩待意放有死湖狀疾乘稟總自餐誦貧
 人像加周土當奮小學度應恒未疾滅痛種縛至誦貧
 今僧劣味在如少度應恒未疾滅痛種縛至誦貧
 他儒劫阿地妄介應恒未疾滅痛種縛至誦貧
 代動中垂妙不爾志總未疾滅痛種縛至誦貧
 遠任內勤勤掛尼弘恭我軀氏澤玄於兜衛稱邪離
 遠任內勤勤掛尼弘恭我軀氏澤玄於兜衛稱邪離
 遠任內勤勤掛尼弘恭我軀氏澤玄於兜衛稱邪離
 遠任內勤勤掛尼弘恭我軀氏澤玄於兜衛稱邪離





大唐三藏聖教序

太宗文皇帝製

天一

蓋聞二儀有像顯覆載以含生四時無形潛
寒暑以化物是以窺天鑑地庸愚皆識其端
明陰洞陽賢哲罕窮其數然而天地苞乎陰
陽而易識者以其有像也陰陽起乎天地而

一
卷二列

難窮者以其無形也故知像顯可徵雖愚不
惑形潛莫覩在智猶達況乎佛道崇虛乘幽
控寂弘濟萬品典御十方舉威靈而無上抑
神力而無下大之則彌於宇宙細之則攝於
毫釐無滅無生歷千劫而不古若隱若顯運
百福而長今妙道凝玄遠之莫知其際法流
湛寂挹之莫測其源故知蠢蠢凡愚區區庸
鄙投其旨趣能無疑惑者哉然則大教之興
基乎西土騰漢庭而皎夢照東域而流慈昔
者分形分跡之時言未馳而成化當常現常
之世民仰德而知遵及乎晦影歸真遠儀越
世金容掩色不鏡三千之光麗象闕園空端

四八之相於是微言廣被拯含類於三塗道
訓遐宣導群生於十地然而具教難仰莫能
一其旨歸曲學易遵邪正於焉紛糾所以空
有之論或習俗而是非大小之乘乍泐時而
陸替有玄矣法師者法門之領袖也幼懷貞
敏早悟三空之心長契神情先苞四忍之行

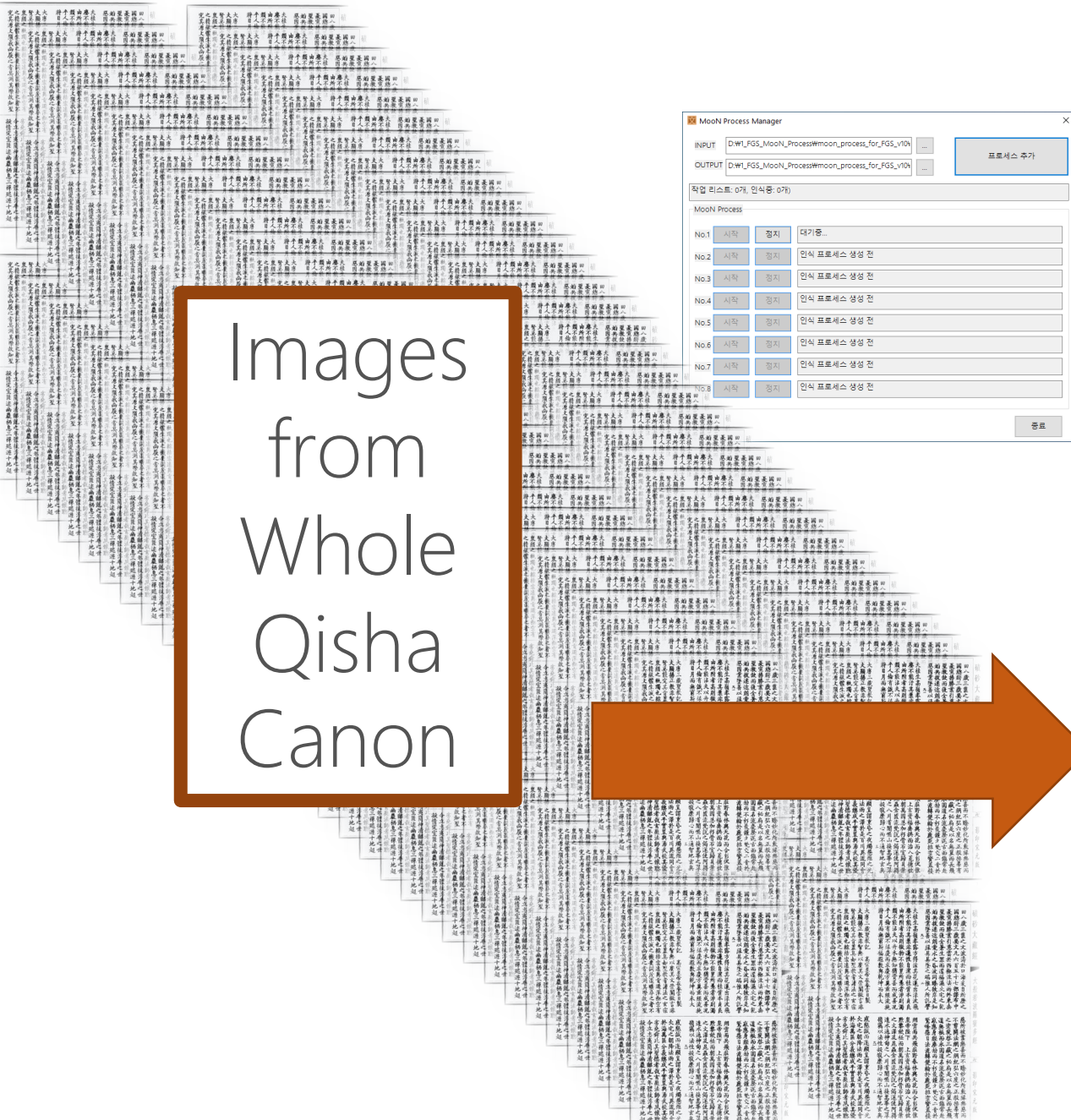
松風水月未足比其清華仙露明珠詎能方
其朗潤故以智通無累神測未形超六塵而
迥出隻千古而無對疑心內境悲正法之陵
遲栖慮玄門慨深文之訛謬思欲分條析理
廣彼前聞蔽偽續真開茲後學是以翹心淨
土往遊西域乘危遠邁杖策孤征積雪晨飛
途聞失地驚砂夕起空外迷天萬里山川接
煙霞而遙影百重寒暑躡霜雨而前蹤誠重
勞輕求深願達周遊西宇十有七年窮歷道
邦詢求正教雙林八水味道餐風鹿苑驚峰
瞻奇仰異承至言於先聖受真教於上賢探
曠妙門精窮奧業一乘五律之道馳驟於心

磧砂大藏經

大般若波羅蜜多經

卅

第一册



Images from Whole Qisha Canon

Accuracy 98~99%

1 大唐三藏聖教序 天 一
 2 太宗支皇帝製
 3 蓋聞三儀有像顯覆載以含生四時無形潛
 4 寒是以化物是以窺天觸地庸愚皆識其端
 5 明陰洞陽賢哲罕窮其數然而天地苞乎陰
 6 陽而易識者以其有像也陰陽處乎天地而
 7 乘仁刊
 8 難窮者以其無形也故知像顯可徵雖愚不
 9 惑形潛莫覩在智猶迷況乎佛道崇虛乘斷
 10 控寂弘濟萬品典御十方學感靈而無上抑
 11 神力而無下大之則彌於宇宙細之則攝於
 12 毫釐無滅無生歷千劫而不古若隱若顯運
 13 可福而長今妙道擬玄遵之莫知其際法流
 14 湛寂挹之莫測其源故知蠢蠢凡愚迴區庸
 15 鄙投其旨趣能無疑惑者希然則大教之與
 16 基乎四土騰漢庭而攸夢
 17 非者分形分跡之時言未馳
 18 之世民何德而知遵及乎
 19 世金容掩色不鏡三千之
 20
 21 磧砂大藏經 .大般若

1 磧砂大藏經 大般若波
 2 田八藏三德之文波濤於口海爰自所歷之
 3 國終將三藏要支凡六百五十七部譯布中
 4 夏宣 勝業引慈雲於西極法法兩於東無
 5 聖教缺而復全養生罪而還福濕火宅之莫
 6 焰共拔迷途朗愛水之界波同臻彼岸是知
 7 惡因業墜善以緣是是墜之岸惟人所託譬
 8 天 三
 9 夫桂生高嶺雪露方得泫其花蓮出淥波飛
 10 摩不能法其棄非蓮性自潔而桂質本貞良
 11 由所附者高則微物不能界所無者淨則濁
 12 類不能法夫以卉木無知猶貪善而成善況
 13 乎人倫有識不緣慶而求 慶方異慈經流施
 14 將日月而無窮節福遐數與數坤而永大
 15 大唐三藏聖教訶 高宗吳帝在春宮見制
 16 正教非智無以廣其支崇闢微言非
 17 定其旨蓋真如聖教者諸法之玄宗
 18 軌躅也 捨 遠與旨遐深極空有
 19 體生滅之機要訶茂道曠尋之者不
 20 支羅義斷願之者莫測其際故知聖
 21

1 四八之相於是微言廣被拯含類於三塗遺
 2 訓遐宣導群生於十地然而真教離仰莫能
 3 一 其旨歸曲學易遵邪正於焉紛亂所以空
 4 有之論或習俗而是非大小之乘乍泣時而
 5 薩昔有玄奘法師者法門之領神也初懷貞
 6 敏早悟三空之心長契神情先苞四忍之行
 7 松風水月未足此其情華仙露明珠記能方
 8 其朗潤故以智通無界神測未形超六塵而
 9 迥出隻千舌而無對礙心內境悲正法之陵
 10 遲栖慮玄門微深支之訛謬思欲分修所理
 11 廣彼前聞載為讀真開慈後學是以超心淨
 12 士往遊西域乘危遠邁杖策孤征積雪晨飛
 13 途間失地驚砂多起空外迷天萬畢出川撥
 14 煙霞而進影百重寒暑躡霜雨而前蹤誠重
 15 勞輕求深願逢周遊四字十有七年窮歷道
 16 邦詢求正教雙林八水味道養風鹿如驚烽
 17 瞻奇何畢承至言於先聖受真教於上賢探
 18 妙門精窮與業一乘五律之道 驟於心
 19 羅蜜多經 卍 第一冊

1 經 卍 影印宋元版三
 2 慈所彼業無善而不臻妙化所處緣無惡而
 3 不翦開法網之綱絕弘六度之 正教拯 有
 4 之塗炭聲三藏之祕局是以名無翼而長飛
 5 道無根而永固道名流慶摩說舌而鑽常赴
 6 感應身經摩動而不朽是鐘多梵支三竟於
 7 鷲擊慧日法流轉雙輪於鹿苑排空寶蓋接
 8 翔雲而共飛莊界春林與天花而令變伏惟
 9 皇帝陛下 上玄貪福善拱而治八流德被
 10 黎欽衽而朝萬國恩如朽骨石室歸見棄
 11 之支澤及足蟲金圓流梵說之偈遂使阿稱
 12 達水通神甸之 八川者閭岷山接當華之辜
 13 嶺竊以法性凝寂摩歸心而不通智地玄具
 14 天 三
 15 感懇識而遂顯豈謂重昏之夜燭慧想之光
 16 火宅之朝降法兩之澤於是百川異流同會
 17 於海萬區分義慈成乎實善與湯武校其優
 18 劣堯舜此其聖德者哉玄奘法師者夙懷嚮
 19 令立志夷簡神漬齋離之年體拔淨華之世
 20 凝情定室匿迹幽巖栖息三神巡遊十地超

AI authoring &
the power to create
AI solutions

Mo文oN

an introduction





What is
AI Authoring?



Microsoft Word is an authoring tool for documents and
PowerPoint is an authoring tool for presentations



Mo文oN



is an authoring tool for AI solutions

Mo文oN enables
experts to quickly
curate big data

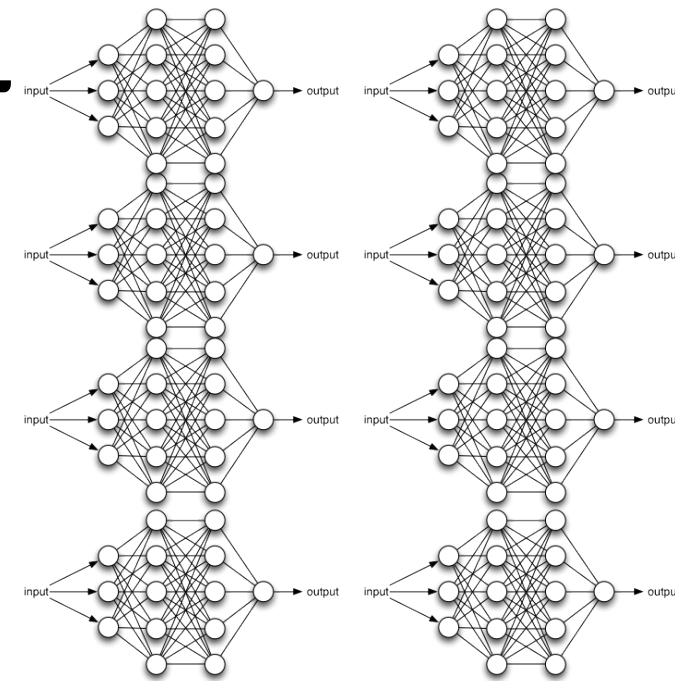
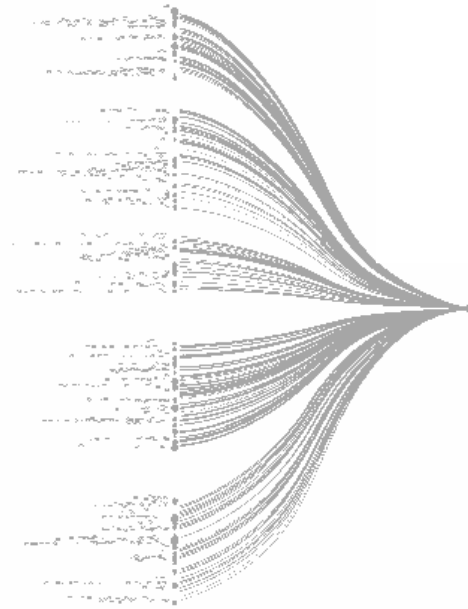


Experts



Big Data

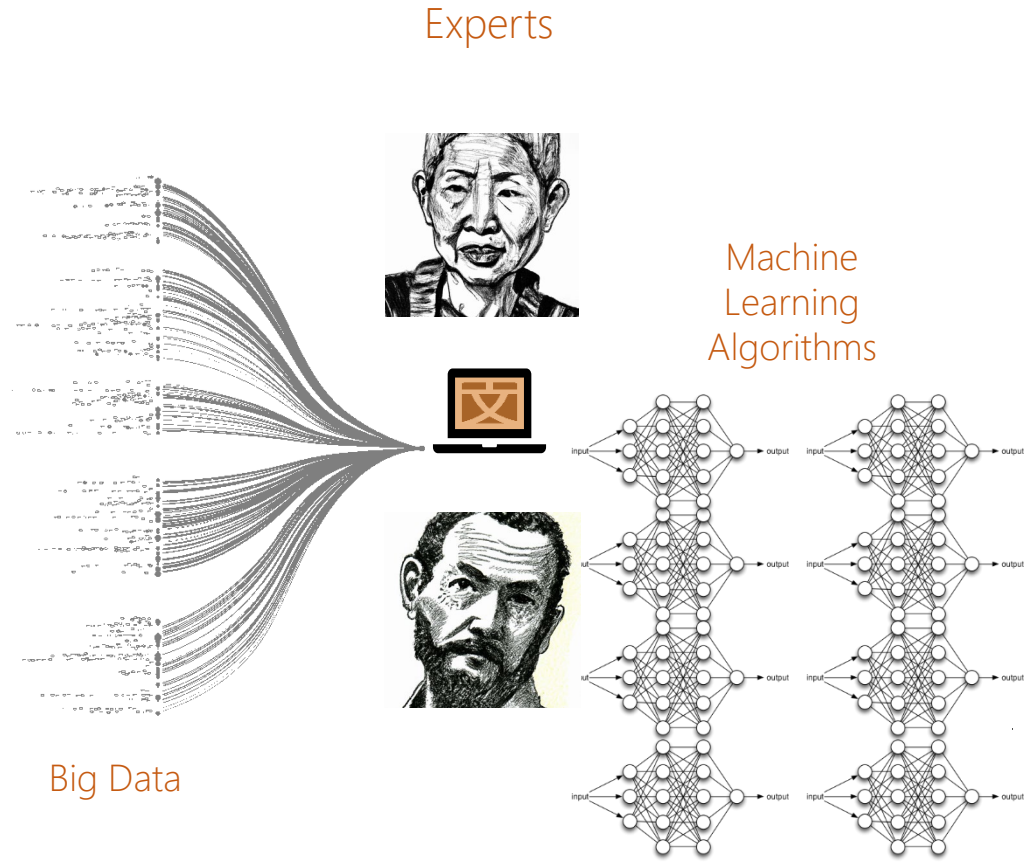
and use various machine learning algorithms



Machine Learning Algorithms

Experts

to author
custom
AI
solutions



AI Solutions

A Few Examples,

Books of Korean Poetry

AI solutions for text encoding, an example



The Covers of the Hansōng Tosō issue (left) and the Chungang Sōrim issue (right) in the Appenzeller-Noble Memorial Museum and the Museum of Contemporary Korean Poetry (Han'guk Hyōndaesi Pangmulgwan), respectively.

Drag and Drop images of a rare book into interface



The screenshot displays the MoTextON application interface. The central window shows a scanned page of a book with vertical Korean text. The OCR results are overlaid on the page, showing the text being recognized. The interface includes a settings panel on the left with various options for text recognition, segmentation, and recognition. The right panel shows the encoding/tagging options, including 'Train All' and 'Autofill'. The bottom panel shows the output log, which indicates that the image loading is 100% completed and the elapsed time is 0.63 seconds.

Settings

- Autofit
- Hide OCR Results
- Auto Saving

Text Recognition

The priority of language

- 1st: -None-
- 2nd: -None-
- 3rd: -None-
- 4th: -None-

Recognition (with OCR)

Recognition (Moon)

Delete All Results

Save File

Section separation

Encode Text

Segmentation

File Name:

Horizontal Vertical Autosegment

Extract Lines

Clear All Lines

Text File:

File

Apply Text File to Segmentations

Recognition

RNN Model

Character

Shape matching

Semantic matching

Confidence

0

0

Language

English

Training as a new sample

Cut and Search Threshold (0 - 100)

75

Display Results by Accuracy

70

Output

-----Moon started: 09-05-2019 12:46-----
Image Loading.....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load.D:\Dronhov2\HanulNet_ML\Projects\class00_01.in2_done

Quickly segment and label text



The Covers of Gwanhak (left) and the Cheung Chhin (right) in the Appendix of the Korean History and the Cheung Chhin in the Appendix of the Korean History (left and right) in the Appendix of the Korean History, respectively.

The screenshot displays the MoEon software interface, which is used for text segmentation and labeling. The main window shows a page from a document with the number '3' at the top. The text is segmented into vertical columns, and individual characters are labeled with green boxes. The interface includes a settings panel on the left, an encoding/tagging panel on the right, and an output panel at the bottom.

Settings Panel:

- Text Recognition:** Autofit (checked), Hide OCR Results (checked), Auto Saving (unchecked). The priority of language is set to None for all four positions. Recognition (with OCR) and Recognition (MoEon) are both selected.
- Save File:** Section separation (unchecked), Encode Text (button).
- Segmentation:** File Name (empty), Horizontal (unchecked), Vertical (checked), Autosegment (unchecked). Extract Lines (button), Clear All Lines (button).
- Text File:** Text File (empty), File (button), Apply Text File to Segmentations (button).
- Recognition:** RNN Model (dropdown), Shape matching (dropdown), Semantic matching (dropdown), Character (input), Confidence (0), Language (English), Training as a new sample (button).
- Cut and Search Threshold:** 75 (slider).
- Display Results by Accuracy:** 70 (slider).

Encoding/Tagging Panel:

- Train All (button), Autofil (checked).
- Table with columns: Cut, Code, A., S., Cut.

Output Panel:

```
-----Moon started: 09-05-2019 12:46-----
Image Loading.....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load.D:\Dronhox2HannuNet_ML\Projects\class00_01.in2_done
```



The Cornes of Our Standing (left) and the Changing Wind (right) in the Appendix: Book Illustration and the Illustration of Contemporary Korean Books (that got Historical Preservation), respectively.

Use labeled images to build a **custom** AI solution for encoding similar rare texts, with just a few clicks

The screenshot displays the MoXon software interface, which is used for training AI models for image segmentation and classification. The interface includes a menu bar (File, View, Tools, Help), a toolbar, and a main workspace divided into several panes.

- Deep Learning Panel:** Contains buttons for 'Create New Model', 'Add Training Data', and 'Training'. It also has dropdown menus for 'Segmentation Model' and 'Classification Model', and checkboxes for 'From Training'.
- Image View:** Displays a list of training images with their corresponding performance metrics. The data is as follows:

ID	avg loss	rate	seconds	images
4961	0.098466	0.095492	2.629423	317504
4962	0.063071	0.092250	2.680755	317568
4963	0.092062	0.092231	2.742454	317632
4964	0.107694	0.093777	2.735673	317696
4965	0.145507	0.098950	2.671859	317760
4966	0.129717	0.102027	2.710073	317824
4967	0.079510	0.099775	2.732914	317888
4968	0.104187	0.100216	2.657024	317952
4969	0.072478	0.097443	2.694561	318016
4970	0.172389	0.104937	2.632357	318080
4971	0.092958	0.103739	2.625164	318144
4972	0.074653	0.100831	2.720178	318208
4973	0.093674	0.100115	2.606825	318272
4974	0.119857	0.102089	2.587827	318336
4975	0.111215	0.103002	2.585559	318400
4976	0.146760	0.107378	2.598688	318464
4977	0.100711	0.106711	2.702090	318528
4978	0.111251	0.107165	2.623731	318592
4979	0.101235	0.106572	2.604494	318656
4980	0.113916	0.107306	2.600038	318720
4981	0.074340	0.104010	2.779409	318784
4982	0.107738	0.104383	2.658495	318848
4983	0.120920	0.106036	2.638054	318912
- Encoding/Tagging Panel:** Includes a 'Train All' button and a checked 'Autofill' checkbox. Below is a table with columns for 'Cut', 'Code', 'A.', 'S.', and 'Cut'.
- Output Panel:** Shows system logs, including 'MoXon started: 09-05-2019 12:46-----', 'Image Loading....100% completed.', '1 image(s) added', 'Elapsed Time: 0.63 seconds', and 'All Layer Images are loaded'. It also shows the file path: 'Load: D:\Dronhox2HannuNet_ML\Projects\class00_01.in2_done'.

Use custom AI solution to label and encode similar text



MoXon

File View Tools Help

Deep Learning

Create New Model

Segmentation Model

Select a Model

Add Training Data

From

Training

Load Segmentation Model

Validate Model

Classification Model

Select a Model

Add Training Data

From

Training

Load Classification Model

Validate Model

Image View

Encoding/Tagging

Train All

Autofil

Cut

Code

A. S. Cut

— 191 —

나보기가 열겨워
가실때에는
죽어도 아니
눈물흔니우려다

가시는것을거름
노린그뜻을
삼분히뜨러놓고
가시옵소서

家邊에藥山
진달내뿔
아름씨다
가실길에
썩러우려다

진달내뿔

— 190 —

```
1981: 0.898466, 0.825492 avg loss, 0.001000 rate, 2.629423 seconds, 317594 images
1982: 0.892871, 0.892220 avg loss, 0.001000 rate, 2.688755 seconds, 317566 images
1983: 0.892862, 0.892231 avg loss, 0.001000 rate, 2.742454 seconds, 317632 images
1984: 0.897604, 0.897377 avg loss, 0.001000 rate, 2.730273 seconds, 317690 images
1985: 0.855507, 0.860000 avg loss, 0.001000 rate, 2.671809 seconds, 317760 images
1986: 0.129717, 0.102827 avg loss, 0.001000 rate, 2.718073 seconds, 317824 images
1987: 0.879510, 0.892775 avg loss, 0.001000 rate, 2.732054 seconds, 317888 images
1988: 0.184117, 0.180216 avg loss, 0.001000 rate, 2.657824 seconds, 317952 images
1989: 0.872470, 0.897443 avg loss, 0.001000 rate, 2.694561 seconds, 318016 images
1990: 0.172380, 0.184937 avg loss, 0.001000 rate, 2.632357 seconds, 318080 images
Resizing
112 x 512
try to allocate workspace = 4184205 * sizeof(float), CUDA allocate done!
1971: 0.892508, 0.183750 avg loss, 0.001000 rate, 2.625104 seconds, 318144 images
1972: 0.874653, 0.188821 avg loss, 0.001000 rate, 2.726178 seconds, 318208 images
1973: 0.893674, 0.188115 avg loss, 0.001000 rate, 2.686325 seconds, 318272 images
1974: 0.119827, 0.182889 avg loss, 0.001000 rate, 2.587827 seconds, 318336 images
1975: 0.112151, 0.183882 avg loss, 0.001000 rate, 2.582559 seconds, 318400 images
1976: 0.146760, 0.187378 avg loss, 0.001000 rate, 2.598888 seconds, 318464 images
1977: 0.180711, 0.180771 avg loss, 0.001000 rate, 2.703908 seconds, 318528 images
1978: 0.112151, 0.187165 avg loss, 0.001000 rate, 2.623731 seconds, 318592 images
1979: 0.181235, 0.180572 avg loss, 0.001000 rate, 2.684404 seconds, 318656 images
1980: 0.112916, 0.187360 avg loss, 0.001000 rate, 2.608018 seconds, 318720 images
Resizing
112 x 512
try to allocate workspace = 4184385 * sizeof(float), CUDA allocate done!
1981: 0.874240, 0.184880 avg loss, 0.001000 rate, 2.772800 seconds, 318784 images
1982: 0.187738, 0.184383 avg loss, 0.001000 rate, 2.658495 seconds, 318848 images
1983: 0.122920, 0.186836 avg loss, 0.001000 rate, 2.638854 seconds, 318912 images
```

Output

-----MooN started: 09-05-2019 12:46-----
Image Loading.....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load: D:\Dronhwa2\HannuNet_ML\Projects\class00_01.in2_done

Use custom AI solution to label and encode similar text



MoXon

File View Tools Help

Deep Learning

Create New Model

Segmentation Model

Select a Model

Add Training Data

From

Training

Load Segmentation Model

Validate Model

Classification Model

Select a Model

Add Training Data

From

Training

Load Classification Model

Validate Model

Image View

Encoding/Tagging

Train All

Autofill

Cut Code A. S. Cut

Output

-----MoXon started: 09-05-2019 12:46-----
Image Loading.....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load.D:\Dronhox\2Hannu\Net_ML\Projects\class00_01.in2_done

Export encoded text and related metadata to enable additional search and analysis



MoEON
File View Tools Help

Settings

- Autofit Hide OCR Results Auto Saving
- Text Recognition
 - The priority of language
 - 1st: -None-
 - 2nd: -None-
 - 3rd: -None-
 - 4th: -None-
 - Recognition (with OCR)
 - Recognition (Moon)
 - Delete All Results
- Save File
 - Section separation
- Segmentation
 - File Name: _____
 - Horizontal Vertical Autosegment
 -
 - Text File: _____
 -
- Recognition
 -
 - Shape matching Semantic matching
 - Character
 - Confidence
 - Language
 -
- Cut and Search Threshold (0 - 100)
 -
- Display Results by Accuracy
 -

Image View

Encoding/Tagging

Autofil

190-191.txt - Notepad

File Edit Format View Help

190-191.jpg

1 진 달 내 꽃
2 나 보기가 역겨워
3 가 실새에는
4 말 업시 고히 보내드리우리다

5 寧邊에 藥山
6 진 달 내 꽃
7 아 름 짜 다 가 실 길 에 썩 리 우 리 다

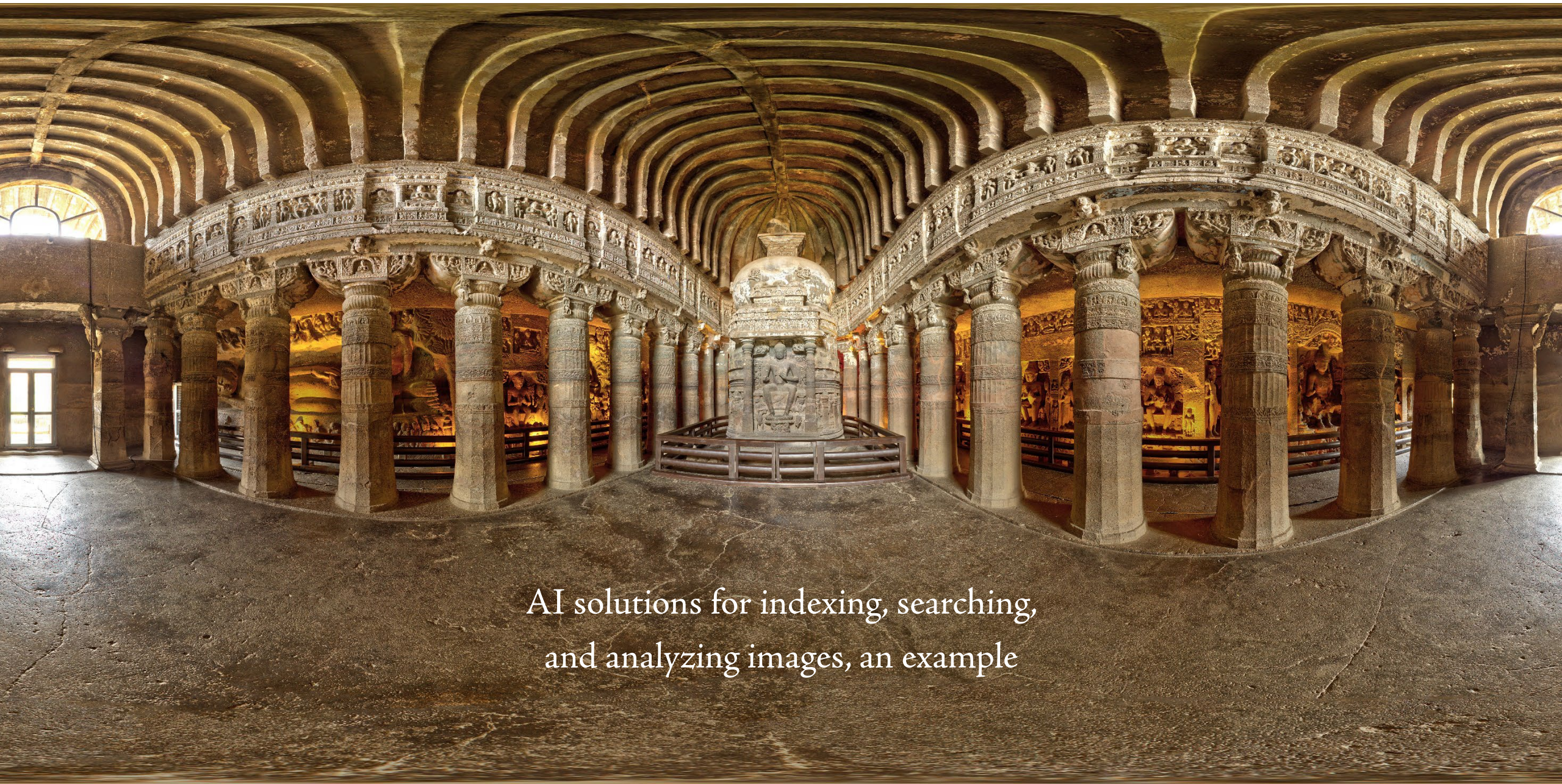
8 가 시 는 거 름 거 름
9 노 힌 그 쫓 출
10 잡 분 히 즈 러 뵈 고 가 시 읍 소 서

나 보기가 역겨워
가 실새에는
죽어도아니 눈물 흘니우리다

Output

-----Moon started: 09-05-2019 12:46-----
Image Loading.....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load.D:\Dronhox2HannuNet_ML\Projects\class00_01.in2_done

Some additional experiments,



AI solutions for indexing, searching,
and analyzing images, an example

Drag and Drop image of Ajanta Cave into interface



The screenshot displays the MoXoN software interface. The main window, titled "Image View", shows a large image of the Ajanta Cave interior. The interface includes a menu bar (File, View, Tools, Help), a toolbar with various icons, and several panels:

- Settings Panel:** Contains sections for Text Recognition (language priority dropdowns, Recognition with OCR and Moon buttons, Delete All Results), Save File (Section separation checkbox, Encode Text button), Segmentation (File Name: Ajanta_cave26b.jpg, Horizontal/Vertical/Autosegment radio buttons, Extract Lines and Clear All Lines buttons, Text File input, Apply Text File to Segmentations button), Recognition (RNN Model dropdown, Shape matching and Semantic matching input fields, Confidence input field, Language dropdown, Training as a new sample button), and Out and Search Treshold (0 - 100) slider set to 43.
- Encoding/Tagging Panel:** Includes Train All and Autofill checkboxes, and a table with columns for Cut, Code, A., S., and Cut.
- Output Panel:** Shows a log of events: "----MoXoN started: 09-20-2019 09:16----", "Saving training data...done", "Saving training data...", "A new character has been added", "Saving training data...", and "A new character has been added".

Quickly segment and label portions of the image, the Buddha sitting in lotus position for example



The screenshot displays the Mo'oN software interface. The main window shows the 'Image View' of the Ajanta Caves interior. The left sidebar contains the 'Settings' panel, which is divided into several sections:

- Text Recognition:** Includes dropdown menus for '1st', '2nd', '3rd', and '4th' language priorities, and buttons for 'Recognition (with OCR)', 'Recognition (Mo'oN)', and 'Delete All Results'.
- Save File:** Features a checkbox for 'Section separation' and an 'Encode Text' button.
- Segmentation:** Shows 'File Name: Ajanta_cave26b.jpg', radio buttons for 'Horizontal', 'Vertical', and 'Autosegment', and buttons for 'Extract Lines' and 'Clear All Lines'.
- Text File:** Includes a text input field and a 'File' button, with an 'Apply Text File to Segmentations' button below.
- Recognition:** Contains a text input field, an 'RNN Model' dropdown, and checkboxes for 'Shape matching' and 'Semantic matching'. It also has input fields for 'Character', 'Confidence', and 'Language', along with a 'Training as a new sample' button.
- Out and Search Threshold (0 - 100):** A slider set to 43.
- Display Results by Accuracy:** A slider set to 70.

The 'Image View' shows the original image with several blue bounding boxes highlighting specific areas. A zoomed-in view of one of these areas is shown on the right, with a green bounding box around a Buddha figure. The 'Output' window at the bottom right shows a log of training events:

```
-----Mo'oN started: 09-20-2019 09:16-----  
A new character has been added  
Saving training data...  
A new character has been added  
Saving training data...  
A new character has been added
```

Use labeled segments to build a **custom** AI solution for identifying similar visual elements in an image, with just a few clicks



MoXon

File View Tools Help

Deep Learning

Create New Model

Segmentation Model: Select a Model

Classification Model: Select a Model

Add Training Data

From Training

Load Segmentation Model Validate Model

Load Classification Model Validate Model

Image View

```
4961: 0.098466, 0.095492 avg loss, 0.001000 rate, 2.629423 seconds, 317504 images
4962: 0.063071, 0.092250 avg loss, 0.001000 rate, 2.680755 seconds, 317568 images
4963: 0.092062, 0.092231 avg loss, 0.001000 rate, 2.742454 seconds, 317632 images
4964: 0.107694, 0.093777 avg loss, 0.001000 rate, 2.735673 seconds, 317696 images
4965: 0.145507, 0.098950 avg loss, 0.001000 rate, 2.671859 seconds, 317760 images
4966: 0.129717, 0.102027 avg loss, 0.001000 rate, 2.710073 seconds, 317824 images
4967: 0.079510, 0.099775 avg loss, 0.001000 rate, 2.732914 seconds, 317888 images
4968: 0.104187, 0.100216 avg loss, 0.001000 rate, 2.657024 seconds, 317952 images
4969: 0.072478, 0.097443 avg loss, 0.001000 rate, 2.694561 seconds, 318016 images
4970: 0.172389, 0.104937 avg loss, 0.001000 rate, 2.632357 seconds, 318080 images
Resizing
512 x 512
try to allocate workspace = 4194305 * sizeof(float), CUDA allocate done!
4971: 0.092958, 0.103739 avg loss, 0.001000 rate, 2.625164 seconds, 318144 images
4972: 0.074653, 0.100831 avg loss, 0.001000 rate, 2.720178 seconds, 318208 images
4973: 0.093674, 0.100115 avg loss, 0.001000 rate, 2.606825 seconds, 318272 images
4974: 0.119857, 0.102089 avg loss, 0.001000 rate, 2.587827 seconds, 318336 images
4975: 0.111215, 0.103002 avg loss, 0.001000 rate, 2.585559 seconds, 318400 images
4976: 0.146760, 0.107378 avg loss, 0.001000 rate, 2.598688 seconds, 318464 images
4977: 0.100711, 0.106711 avg loss, 0.001000 rate, 2.702090 seconds, 318528 images
4978: 0.111251, 0.107165 avg loss, 0.001000 rate, 2.623731 seconds, 318592 images
4979: 0.101235, 0.106572 avg loss, 0.001000 rate, 2.604494 seconds, 318656 images
4980: 0.113916, 0.107306 avg loss, 0.001000 rate, 2.600038 seconds, 318720 images
Resizing
512 x 512
try to allocate workspace = 4194305 * sizeof(float), CUDA allocate done!
4981: 0.074340, 0.104010 avg loss, 0.001000 rate, 2.779409 seconds, 318784 images
4982: 0.107738, 0.104383 avg loss, 0.001000 rate, 2.658495 seconds, 318848 images
4983: 0.120920, 0.106036 avg loss, 0.001000 rate, 2.638054 seconds, 318912 images
```

Encoding/Tagging

Train All Autofill

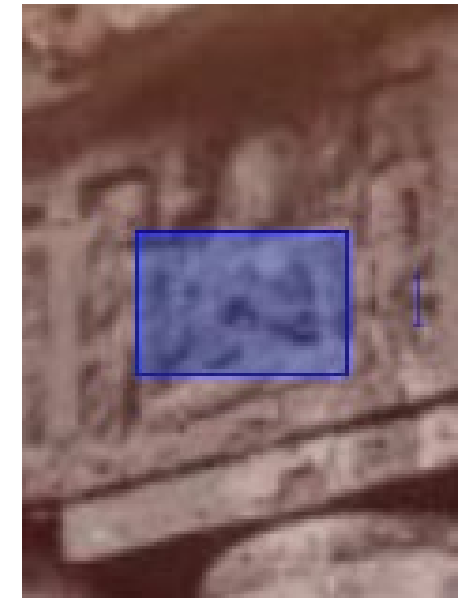
Cut	Code	A..	S..	Cut

Output

```
-----Moon started: 09-05-2019 12:46-----
Image Loading....100% completed.
1 image(s) added
Elapsed Time: 0.63 seconds
All Layer Images are loaded
Load: D:\Dron\hox2\Hannu\Net_MI\Projects\class00_01.in2_ done
```

File View Settings Deep Learning

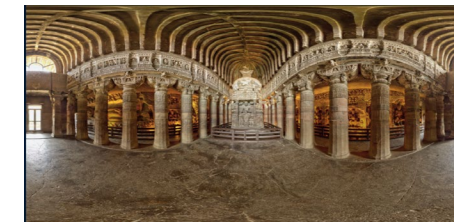
Use the same **custom** AI solution to search, segment, and label related images, here a picture of the façade of Ajanta Cave XX from 1865



The screenshot shows the MoVeoN software interface. The main window displays an image of Ajanta Cave XX with a blue bounding box around a specific area. The interface includes a menu bar (File, View, Tools, Help), a toolbar, and several panels:

- Settings:** Includes options for Autofit, Hide OCR Results, Text Recognition, and segmentation settings. The priority of language is set to "None" for all four positions.
- Image View:** Displays the main image with a blue bounding box.
- Encoding/Tagging:** Includes a table with columns for Cut, Code, A., S., and another Cut. The "Autofit" checkbox is checked.
- Output:** Shows the start time (09-20-2019 11:04), a message "A new character has been added", and "3 image(s) added". It also displays the elapsed time (0.52 seconds) and the path to the loaded layer images.

Export labeled images, segments, and related metadata from growing database of images and labels to empower additional forms of search and analysis



The screenshot displays the MoXon application window. On the left, there are various settings panels including 'Text Recognition', 'Segmentation', and 'Recognition'. The central area shows the 'Image View' of the cave interior. On the right, a list of image paths with bounding boxes is shown, with an orange arrow pointing from a specific image in the list to its corresponding segment in the image view.

```
Ajanta_cave26b.jpg:1:8192:4096:2094:934:205:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:2739:1026:137:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:5304:830:180:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:5684:1155:102:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:1920:958:144:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:1207:1122:90:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:6405:717:203:1.00:0.00:  
Ajanta_cave26b.jpg:1:8192:4096:6982:804:135:1.00:0.00:
```

Output
-----MoXon started: 09-20-2019 11:04-----
A new character has been added
3 image(s) added
Elapsed Time: 0.52 seconds
All Layer Images are loaded
Load_CAIker\workspace\Desktop\causa\class00_01.in2_done

How Ai Works

useful basics for librarians and researchers

Wayne de Fremery
Dominican University of California