

철학은 다양한 학문과 함께 쓰일 수 있다. 대표적인 것으로는 형이상학, 윤리학, 성치철학, 과학 철학, 언어철학, 사회철학, 논리학 등이 있다.

소크라테스 이전은 철학의 원 1 대상은 사언이었다. 이것을 그리스 철학이라고 한다. 사언은 스스로 움직이는 것
었다. 특
테스 이후
대상을 놓

중세시대(신에 대한
나가 대상
경험론을
세기의 철
현대 철학
을 언급한

Korean OCR

SSIP '09, Debrecen

겨났으며
철학과 다른 학문을 구분하는 방법
을 살펴 보는 것이다. 18세기까지만 하더라도 물리학은 철학과
식되었다. 물론 여기서 당시에서 농용된 물리학은 오늘날 자연
있음은 말할 나위가 없다.

철학의 고유한 문제들은 18 세기의 철학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

웬 초콜릿? 제가 원했
던 건 뱅튀기와 의류예
요. 애야, 왜 또 불

Team 3

Hrvoje Kalinić - University of Zagreb

Gabor Nemeth - University of Szeged



Peter Kardos

University of Szeged

Szymon Gajek - Jagiellonian University

Outline

- The problem
- Korean alphabet
- Proposed method
 - Skeletonization
 - Line extraction
 - Syllable extraction
 - Character (jamo) detection
 - Feature extraction
 - Classification
- Results
- Future work

질학은 다양한 학문과 함께 쓰일 수 있다. 대표적인 것으로는 형이상학, 윤리학, 성치질학, 과학 질학, 언어질학, 사회질학, 논리학 등이 있다.

소크라테스 이전은 질학의 연구대상을 자연이었다. 이들을 그리스 철학이라고 한다. 사인은 스스로 움직이는 것이었다. 특히 소크라테스 이후는 질학의 연구대상을 인간으로 바꾸었다. 후반, 즉 소크라테스 때는 질학은 인간의 혼을 대상으로 하였다. 소크라테스는 그리스 철학과 반대하는 생각을 하였다. 소크라테스는 소크라테스의 철학의 대상과 그리스 철학의 대상을 놓고 논쟁을 벌였다.

The problem formulation

- **Optimistic:**
 - For the given image of the Korean text rewrite the text in the romanized version

- **Realistic:**
 - Extract the characters from the text and train the model for character classification

현대 질학은 언어질학과 과학 질학으로 나뉘었다. 언어질학은 언어의 구조와 기능을 연구하는 학문이다. 과학 질학은 자연 현상의 구조와 기능을 연구하는 학문이다. 언어질학과 과학 질학은 서로 다른 학문이지만, 둘 다 언어와 과학을 연구하는 학문이다.

그러나, 18세기까지만 하더라도 물리학은 질학과 다른 학문을 구분하는 방법 중 하나는 철학이 제기하는 문제를 살피는 것이다. 18세기까지만 하더라도 물리학은 질학과 다른 학문을 구분하는 방법 중 하나는 철학이 제기하는 문제를 살피는 것이다. 18세기까지만 하더라도 물리학은 질학과 다른 학문을 구분하는 방법 중 하나는 철학이 제기하는 문제를 살피는 것이다.

질학의 고유한 문제들은 18세기의 질학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

Korean alphabet

7 L C 2 □ H 人 O ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ

대신 중서 신어

ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ

加 여 있

경임루를 네 입리루와 경임루는 동시에 하였다. 이것을

- Phonetic scripture

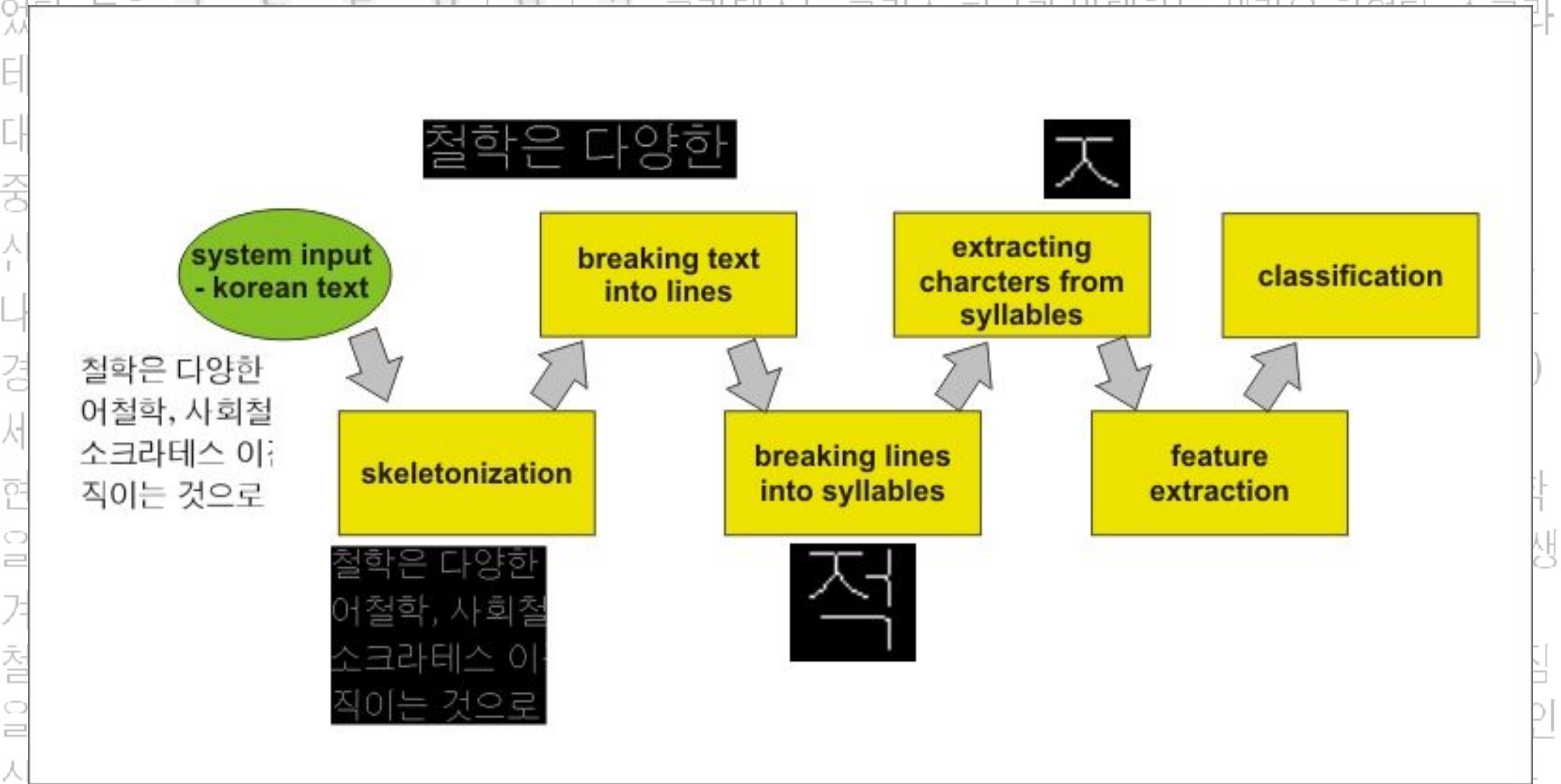
- No jamo can stand alone
- Ligatures and accidentally touching characters

질학의 고유한 문제들은 18 세기의 질학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

철학은 다양한 학문과 함께 쓰일 수 있다. 대표적인 것으로는 형이상학, 윤리학, 성치철학, 과학 철학, 언어철학, 사회철학, 논리학 등이 있다.

소크라테스 이전은 철학의 인 / 대상 / 사인이다. 이것을 근대철학이라고 한다. 사인은 스스로 움직이는 것 / 후반, 즉 소크라테스때는 철학은 인간의 혼을 대상으로 하는 것이다.

Proposed method



요. 해야, 왜 또 물

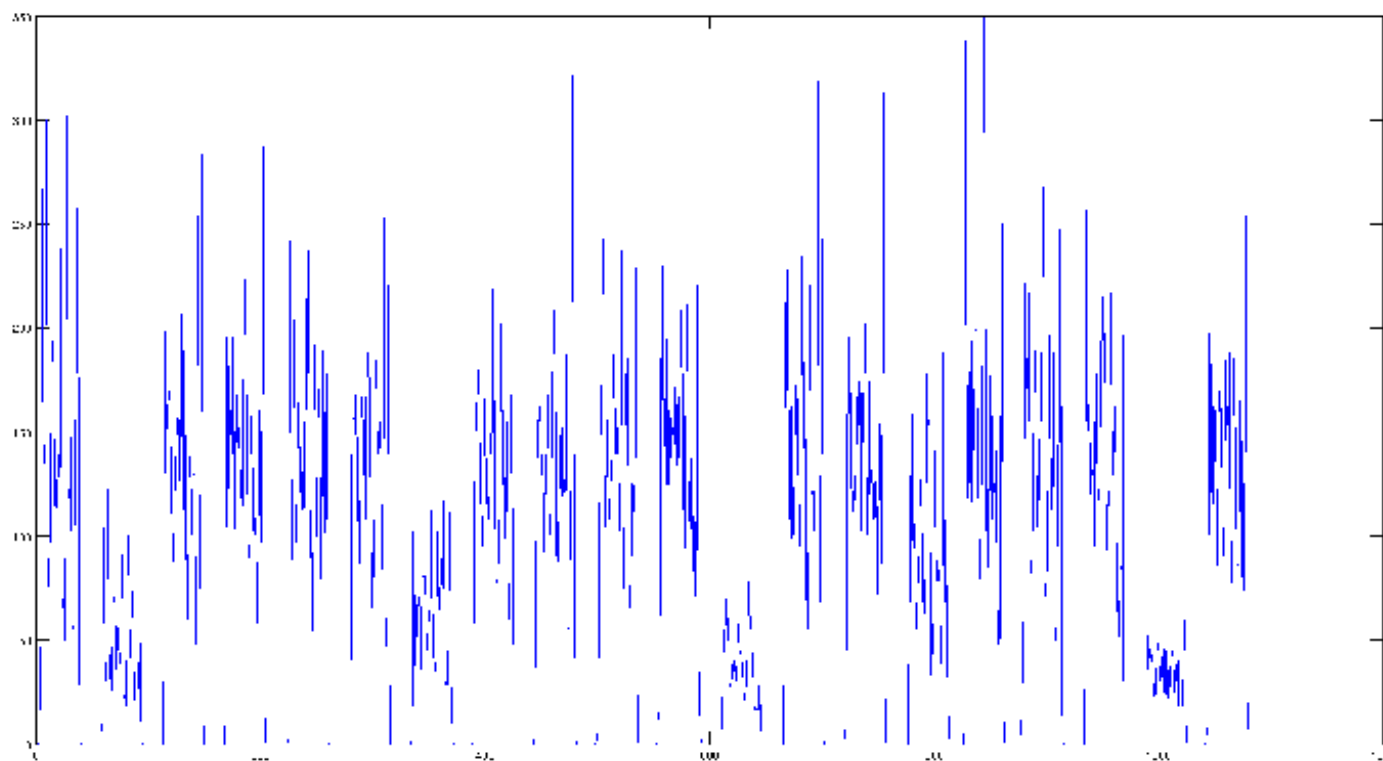
있음은 말할 나위가 없다. 철학의 고유한 문제들은 18 세기의 철학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

철학은 다양한 학문과 함께 쓰일 수 있다. 대표적인 것으로는 형이상학, 윤리학, 정치철학, 과학 철학, 언어철학, 사회철학, 논리학 등이 있다.

소크라테스 이전은 철학의 원시 대상을 사물이었다. 이것을 소크라테스 철학이라고 한다. 사인은 스스로 움직이는 것이었다. 특히 소크라테스 이후 철학은 인간의 혼을 대상으로 한다. 이들은 소크라테스의 철학의 대상과 그리스 철학의 대상을 놓

Line extraction

• Y-axis projection

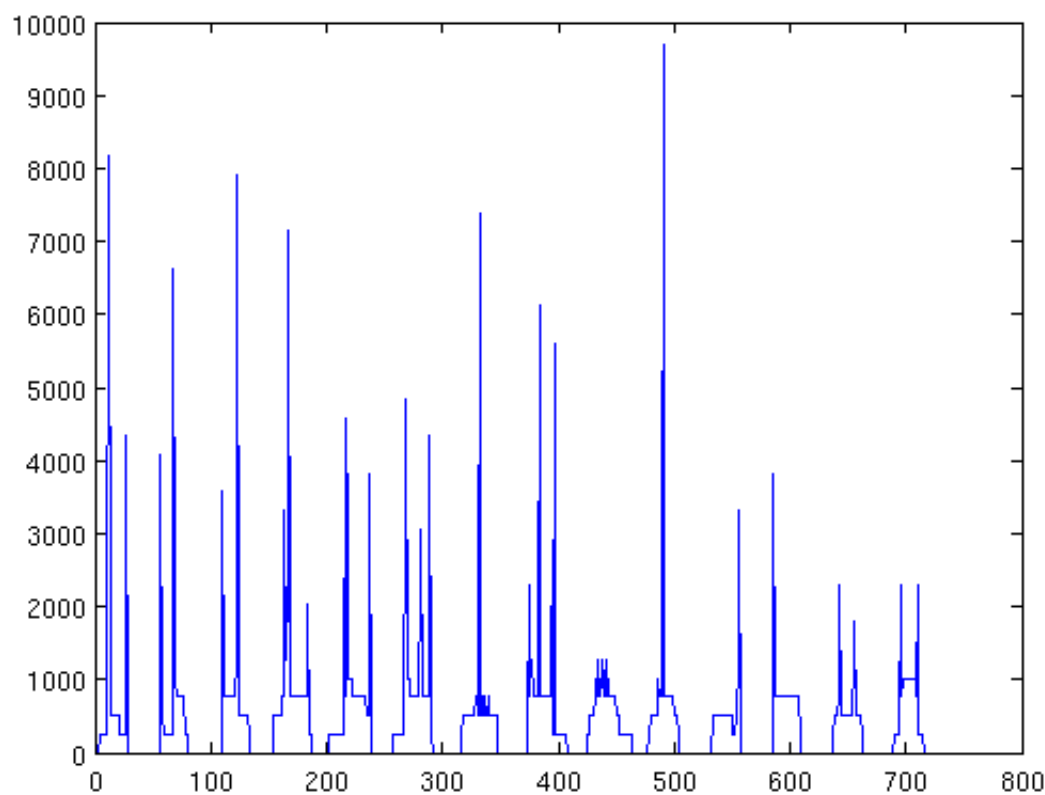


중세시대
신에 대한
나가 대상
경험을
세기의 철
현대 철학
을 언급한
겨났으며
철학과 다
을 살펴 보
식되었다.
있음은 말할 나위가 없다.
철학의 고유한 문제들은 18 세기의 철학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

Syllable extraction

Syllable extraction

- X-axis projection



다 강하게 띠게 되어
리 내로 인식할수 있
제한 것은 입리론과
이것은 19 세기와 20

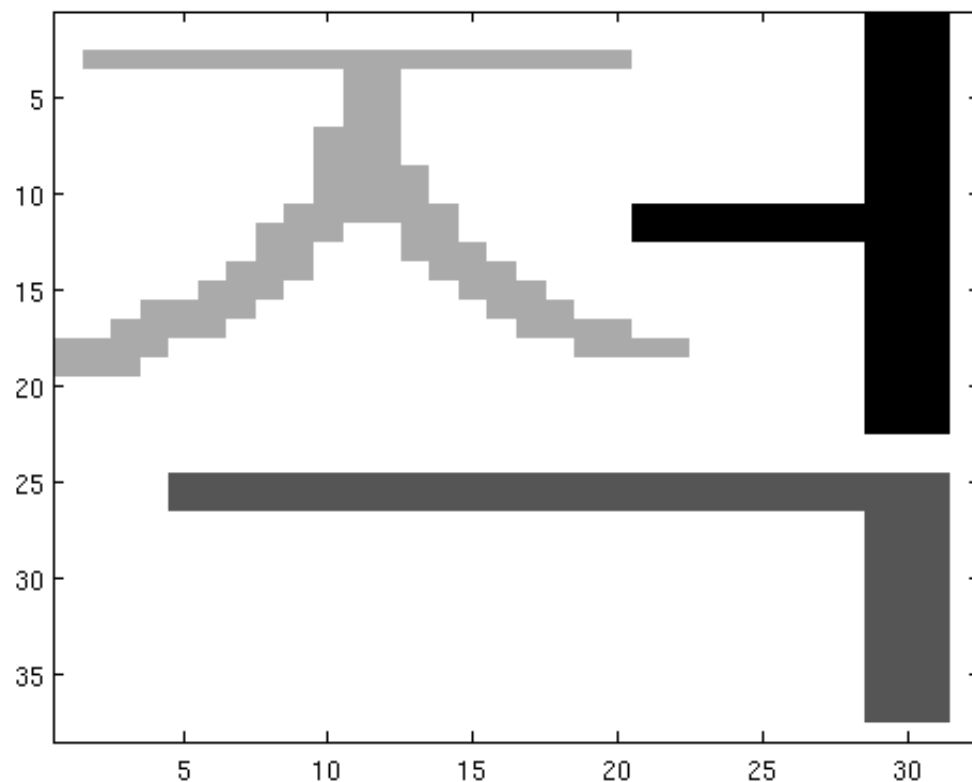
르가 처음 언어 질학과 크게 연관되어 생

테가 원했
와 의류예
또 보
는 짐
루 인
이가

요. 애야, 왜 또

철학의 고유한 문제들은 18 세기의 철학자 임마누엘 칸트의 네 가지 물음으로 요약될 수 있다.

**늘
더
했**



하게 되었고
부 **했** 수 있
것은 입리춘과
은 19 세기와 20

1 처음 언어 질학
크게 연관 되어 생

가 원했
의류에 는 짐
부인

요. 애야, 왜 또

Classification

- Manual selection of characters for the database

- 807 characters

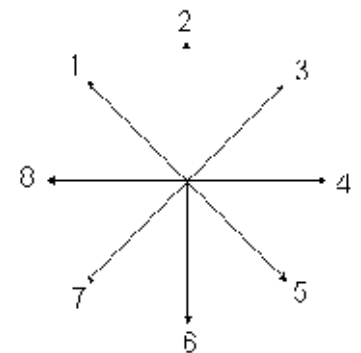
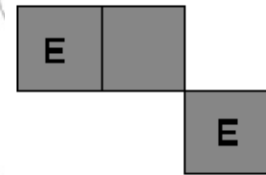
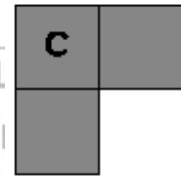
- Feature selection

- Endpoints, junction points, corner points, image size, x and y projections...

- Classification in 24 classes

- R statistical tool used

- Leave N out validation (N = 30% of the dataset)



Results

- Decision tree model constructed

- 87% accuracy

- SVM model constructed

- 98% accuracy

- Random criteria: 4.17%

- Partially accurate character detection from the scanned text

- Text dependent error

Future work

- Character detection can be improved using expert knowledge
 - Algorithm proposed on the web page
- Automatic romanization (transliteration)
- User interface

Korean Alphabet											
Consonants											
ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ
g,k	n	d,t	r,l	m	b,p	s	ng	j	ch	k	t
silent in initial position											
ㄲ	ㄴ	ㄸ	ㄹ	ㅁ	ㅃ	ㅆ	ㅇ	ㅉ	ㅊ	ㅋ	ㅌ
kk	tt	pp	ss	jj							
Vowels											
ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ		
a	ya	eo	y eo	o	yo	u	yu	eu	i		
father	saw	home	moon	put	meet						
ㅐ	ㅒ	ㅖ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ			
ae	yae	e	ye	wa	wae	oe	wo	we	wi	ui	
hand	set					wet					

The end



Gabor Nemeth

University of Szeged

Peter Kardos

University of Szeged

Szymon Gajek

Jagiellonian University

Hrvoje Kalinić

University of Zagreb