

佛教知识图谱及问答

东南大学 认知智能研究所

漆桂林教授

概述-佛学知识图谱

◆首个中文佛教知识图谱

◆KG-Buddhism V1.0

3,152个人物类型实体

136,162条三元组信息

12类属性，15种infobox属性

◆在线API

数据访问地址：<http://www.kg-buddhism.com>



概述-佛学考试机器人

◆依托于863国家重点科研项目-高考机器人技术团队

◆佛学语料

佛学人物知识图谱，佛学术语知识图谱

11个类目佛学词典

3200条佛学百科知识，14部佛学教材

上千条高质量问答对，25部重要佛经的原文-译文对

◆考试成绩

基于搜集到的175道佛教试题进行测试，准确率达**64%**。



目录

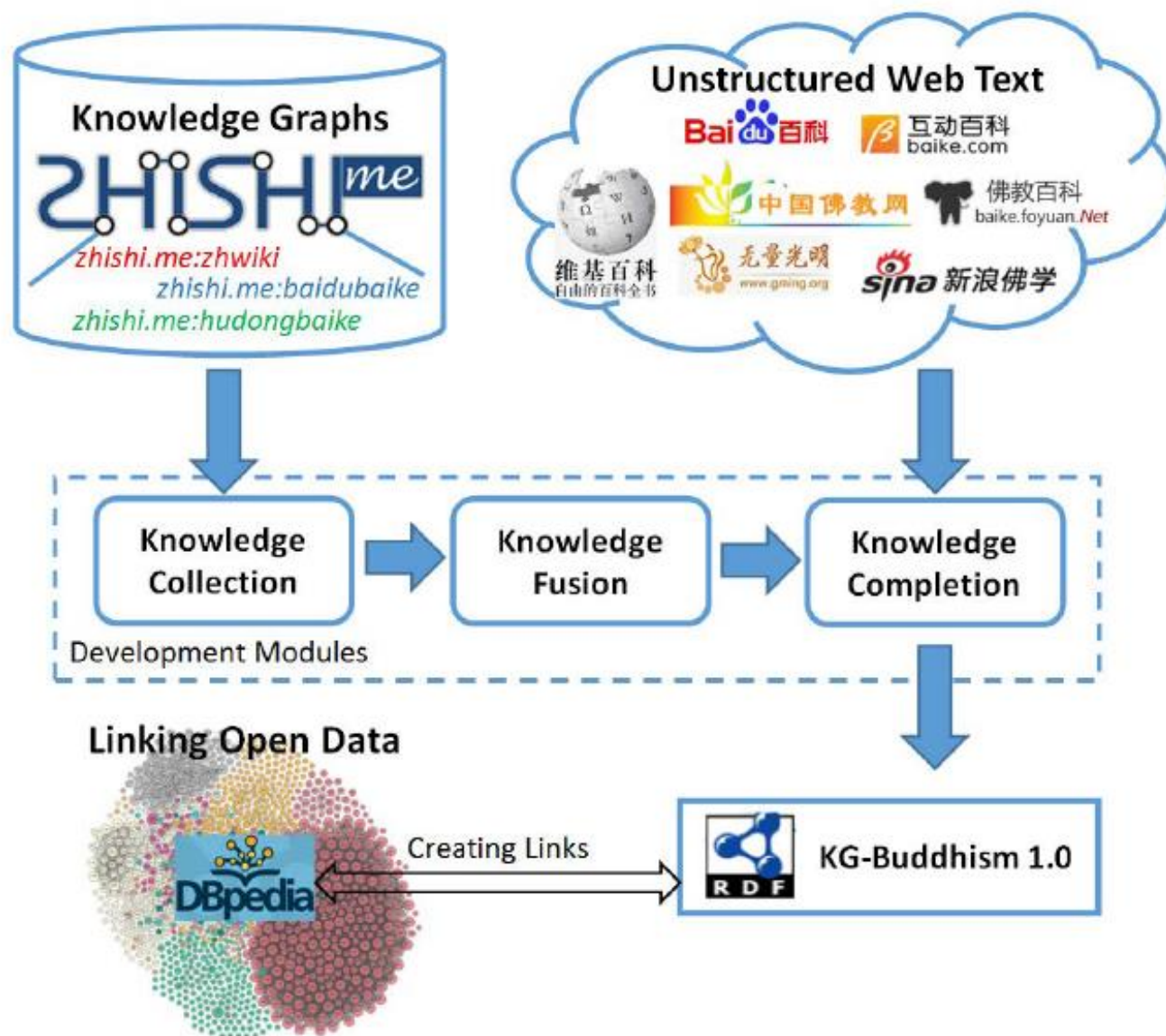
知识图谱

构建方法，数据统计，在线API

考试机器人

题型分析，系统框架，Demo展示

图谱构建框架



知识收集

Category方法

- 观察百科中与佛教人物相关的分类
- 抽取佛教人物分类下所有文章对应的实体

知识收集

分类:佛教

此页面分类的主条目是**佛教**。

子分类

本分类有以下30个子分类，共有30个子分类。

- ▶ [各国佛教](#) (19个分类， 13个页面)
- ▶ [各大洲佛教](#) (3个分类)

*

- ▶ [佛教模板](#) (7个分类， 27个页面)
- ▶ [佛教相關列表](#) (2个分类， 3个页面)

B

- ▶ [佛教小作品](#) (1个分类， 403个页面)

C

- ▶ [佛教宇宙論](#) (2个分类， 9个页面)

佛

- ▶ [佛教之最](#) (1个页面)
- ▶ [佛教修行法門](#) (29个页面)
- ▶ [佛教历史](#) (8个分类， 90个页面)
- ▶ [佛教咒語](#) (15个页面)
- ▶ [佛教哲學](#) (3个分类， 6个页面)
- ▶ [佛教圣者](#) (3个分类)
- ▶ [佛教宗派](#) (10个分类， 11个页面)
- ▶ [佛教建筑](#) (1个分类， 1个页面)
- ▶ [佛教戒律](#) (2个分类， 23个页面)
- ▶ [佛教政治](#) (2个分类， 1个页面)
- ▶ [佛教文化](#) (12个分类， 15个页面)
- ▶ [佛教文学](#) (8个页面)

- ▶ [佛教文献](#) (6个分类， 2个页面)
- ▶ [佛教相關人物](#) (5个分类， 1个页面)
- ▶ [佛教研究](#) (2个分类， 6个页面)
- ▶ [佛教神祇](#) (8个分类， 37个页面)
- ▶ [佛教神話](#) (14个页面)
- ▶ [佛教組織](#) (5个分类， 8个页面)
- ▶ [佛教與其他宗教](#) (2个页面)
- ▶ [佛教與女性](#) (1个分类， 1个页面)
- ▶ [佛教节日](#) (3个分类， 5个页面)
- ▶ [佛教術語](#) (7个分类， 449个页面)
- ▶ [佛教题材作品](#) (5个分类， 12个页面)

反

- ▶ [反佛教](#) (1个分类， 13个页面)

知识收集

Category方法

- 观察百科中与佛教人物相关的分类
- 抽取佛教人物分类下所有文章对应的实体

命名规则方法

e.g. “.*菩萨” “.*禅师”

- 维基百科“佛教头衔”分类下的所有实体
- 已抽取出的实体名中高频的公共字符串

知识融合

主语融合

- 实体的“别名”属性和重定向作为实体的别名集合
- 不同来源的实体存在一个完全匹配的别名则认为是相同实体
- 人工检查相同实体数多于三个的映射

知识融合

主语融合

百度百科 :	<code><http://zhishi.me/hudongbaike/resource/额尔德尼·确吉坚赞></code>	{十世班禅, 班禅额尔德尼·确吉坚赞, 第十世班禅额尔德尼}
互动百科 :	<code><http://zhishi.me/baidubaike/resource/班禅额尔德尼·确吉坚赞></code>	{确吉坚赞, 班禅额尔德尼·确吉坚赞, 罗桑赤烈伦珠}
维基百科 :	<code><http://zhishi.me/zhwiki/resource/十世班禅></code>	{班禅额尔德尼·确吉坚赞, 额尔德尼·确吉坚赞, 罗桑赤烈伦珠确吉坚赞}



`<http://www.kg-Buddhism.com/entity/额尔德尼·确吉坚赞>`

知识融合

谓语融合

- Infobox属性：保留选定的15个子属性，人工总结每个属性在现有知识图谱中存在的谓语形式
- 其它属性：直接替换谓语的命名空间

<http://www.zhishi.me/> ➡ <http://www.kg-buddhism.com/>

知识融合

宾语融合

➤ 单值属性

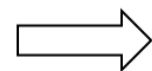
triple 1:
<http://www.kg-Buddhism.com/entity/释星云>
<http://www.kg-Buddhism.com/property/出生日期>
“1927-09-19”@zh

triple 2:
<http://www.kg-Buddhism.com/entity/释星云>
<http://www.kg-Buddhism.com/property/出生日期>
“1927-08-19”@zh

triple 3:
<http://www.kg-Buddhism.com/entity/释星云>
<http://www.kg-Buddhism.com/property/出生日期>
“1927-08-19”@zh

大多数

大多数原则



KG-Buddhism 1.0:
<http://www.kg-Buddhism.com/entity/释星云>
<http://www.kg-Buddhism.com/property/出生日期>
“1927-08-19”@zh

精确性原则：日期、地点等类型的属性值出现冲突时选择最精确的一个。

大多数原则：不同来源的属性值出现冲突时，选择出现次数最多的值。

➤ 多值属性

直接合并去重

知识补全

知识补全

- 对infobox属性进行补全
- 人工编写规则从文本中抽取属性值

e.g. 出生日期，死亡日期： $((\backslash d^+ \text{年}) - (\backslash d^+) \text{年})$

数据集统计

➤ 实体

3,152 个人物类型实体

1,486 条中文DBpedia链接

3,822 条Zhishi.me链接

➤ 知识

136,162 条三元组

12类属性, 15种infobox属性

属性	拥有该属性的实体数	三元组总数
Abstracts	2017	3235
Aliases	180	290
Categories	3149	15216
External Links	1009	3390
Images	1204	4894
Internal Links	2446	34887
Infobox Properties	2632	17540
Labels	3152	3152
Redirects	422	2169
Related Pages	1220	45067
DBpedia Links	1470	1486
Zhishi.me Links	3152	3822

实体覆盖率统计表

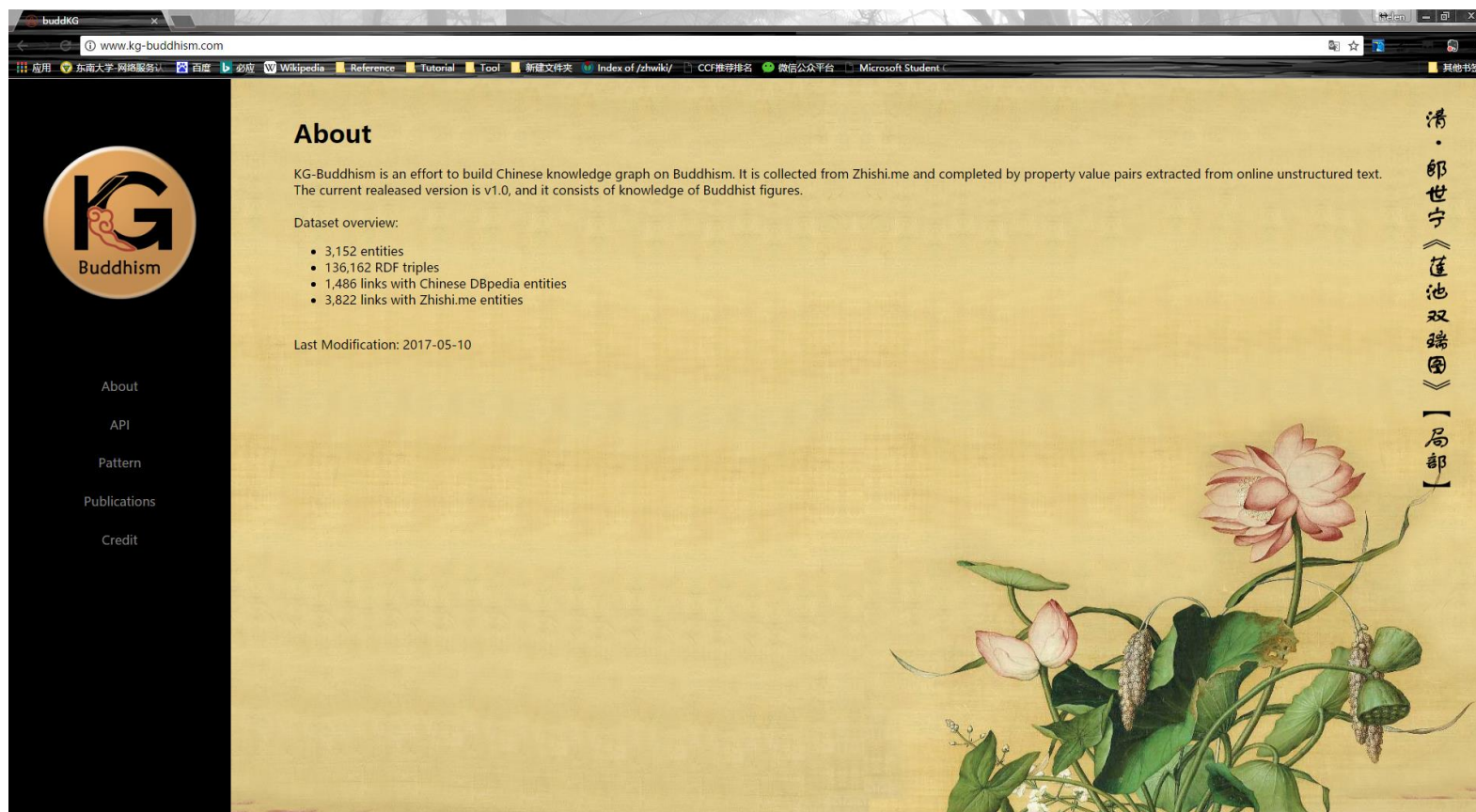
数据集统计

属性	拥有该属性的实体数		属性	拥有该属性的实体数	
	补全前	补全后		补全前	补全后
出生日期	394	936	法名	107	536
死亡日期	319	842	别名	215	1608
出生地	275	634	宗派	126	898
籍贯	139	227	师傅	16	356
代表作	107	211	弟子	10	615
主要成就	113	832	时代	132	132
身份	0	151	国籍	233	1891
俗名	542	1261			

知识补全统计表

在线访问API

➤ 数据访问地址：<http://www.kg-buddhism.com>



调用API

➤ 查询方式

1. 查询给定实体的所有知识：

`www.kg-buddhism.com/cnschema/entities/[label]`

2. 查询给定实体的特定属性值：

`www.kg-buddhism.com/cnschema/entities/[label]&[property]`

➤ 返回格式

Response	Code	Content
Success Response	200	<code>{"@id":"李叔同","s:alternateName":"弘一法师","cns:student":"丰子恺"}</code>
Error Response	404 NOT FOUND	<code>{"error":"Entity doesn't exist"}</code>
	401 UNAUTHORIZED	<code>{"error" : "You are unauthorized to make this request."}</code>

考试机器人

题型分析, 系统框架, Demo展示

考试机器人

考试机器人：Todai Robot Project (2014, Japan)
the Project Aristo (2015, America)

863课题：高考地理 考试机器人

佛教入学考试：一个更具挑战性的任务。

1. 僧伽是①涅槃义②和合众③杀贼义。
2. 「诸行无常、诸法无我、涅槃寂静」称为①三种无常②三法印③三乘道。
3. 人生最大的错误是①杀生②妄语③邪见。



典型问题 (内容/答案类型划分)

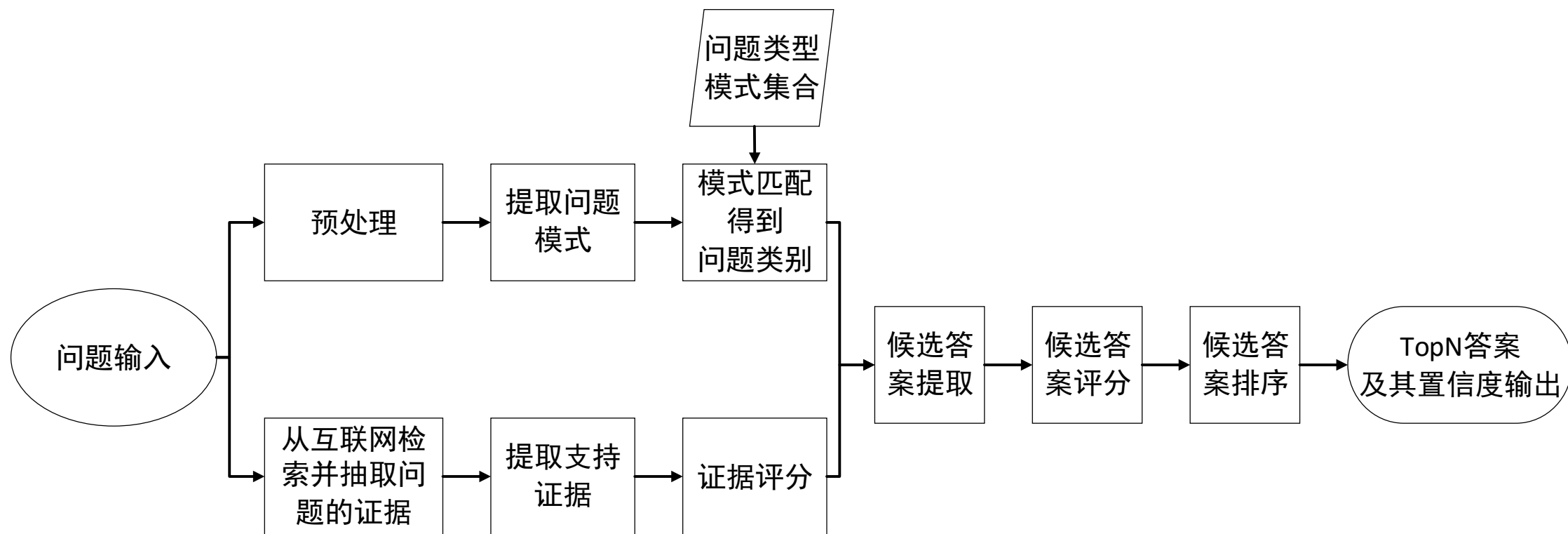
1. 人物类 佛陀在世时，印度最忠诚拥护佛法的国王是()？
2. 地名类 中国佛教四大名山中，观音菩萨的道场是()？
3. 佛经名类 善财童子五十三参出自()？
4. 术语类 十善业又分身、口、意三类，意善业有三，是指()？

典型问题 (提问方式划分)

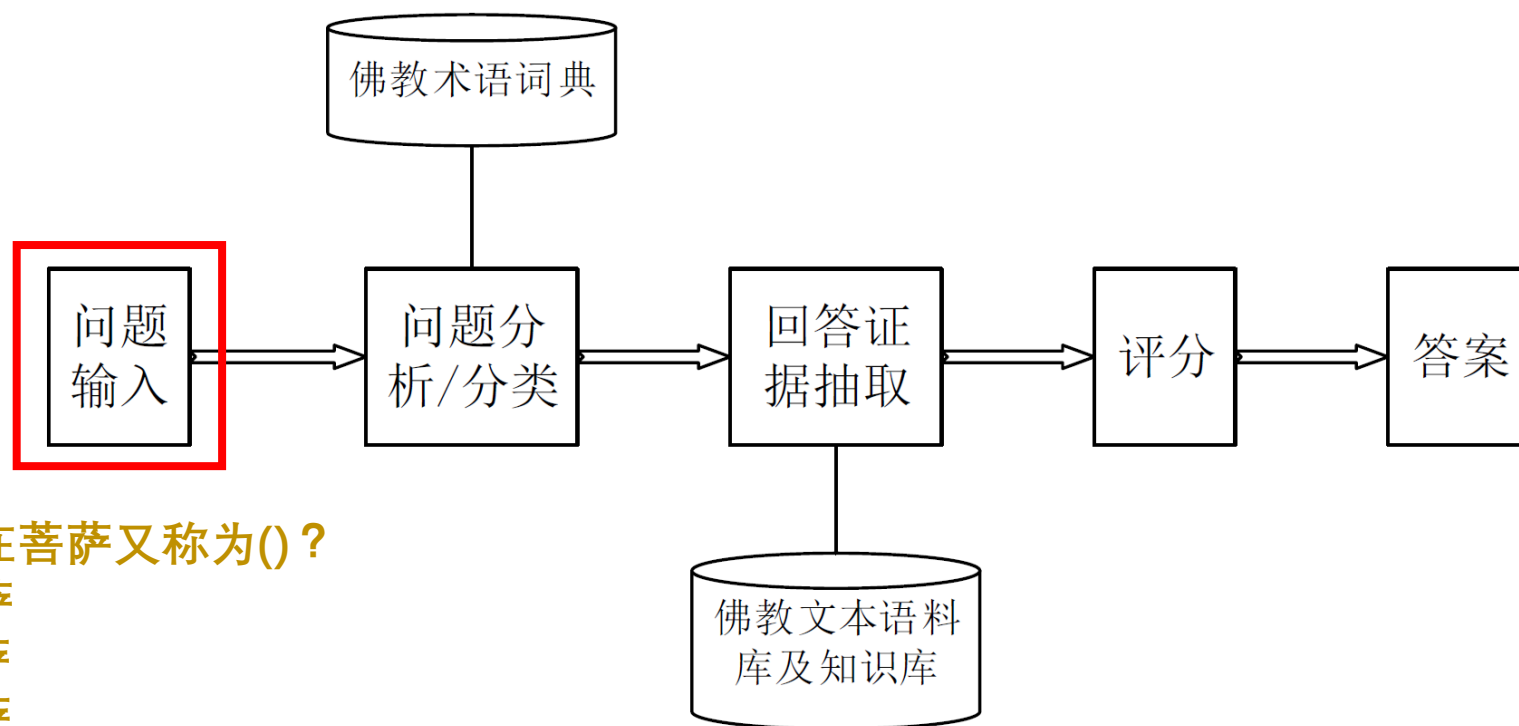
1. 直答类 佛陀在()为五比丘说法？
2. 解释类 佛陀十号之一的两足尊，其含义是指()？
3. 列举类 一般佛教徒经常诵读的经典是()？

以及，选择题中特有的问题句式，如：以下何部经典不是谈般若空性？

框架结构 V1.0



框架结构 V2.0

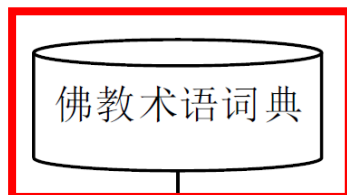


《心经》中的观自在菩萨又称为()？

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨

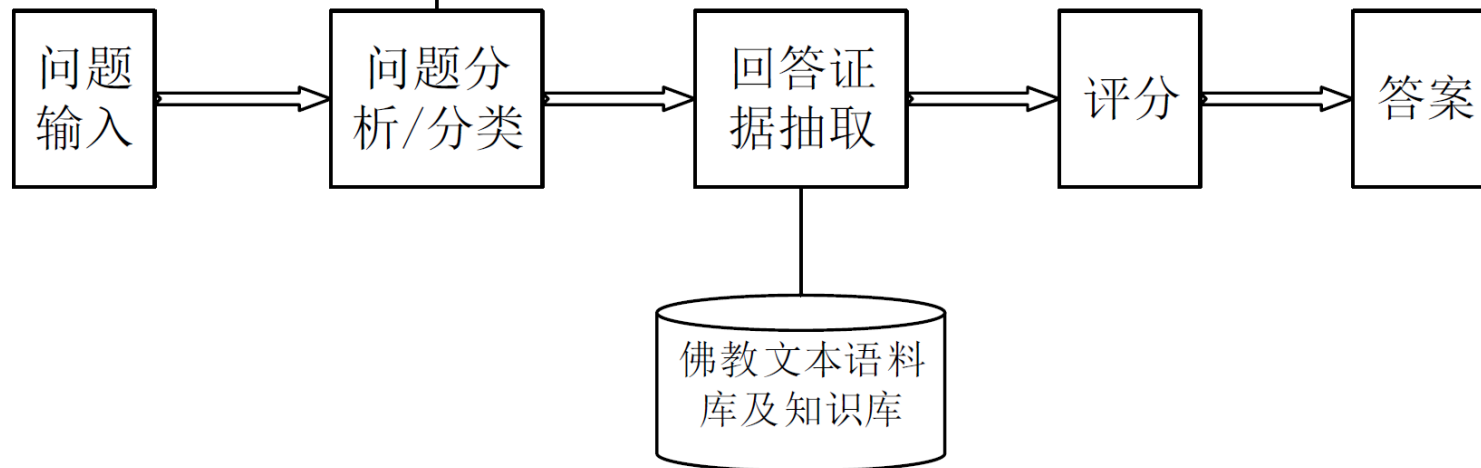
框架结构 V2.0

佛学人物词典/书籍词典等



《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨

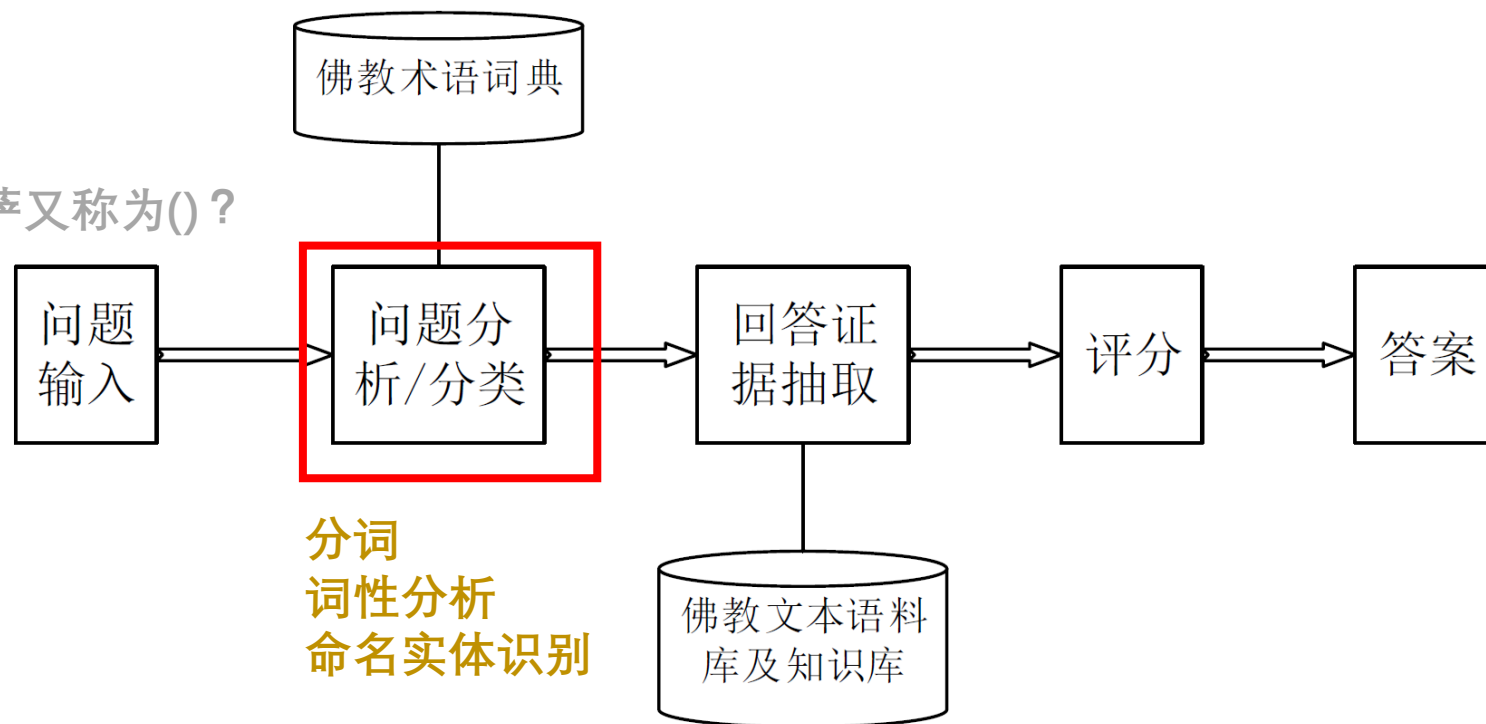


框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨

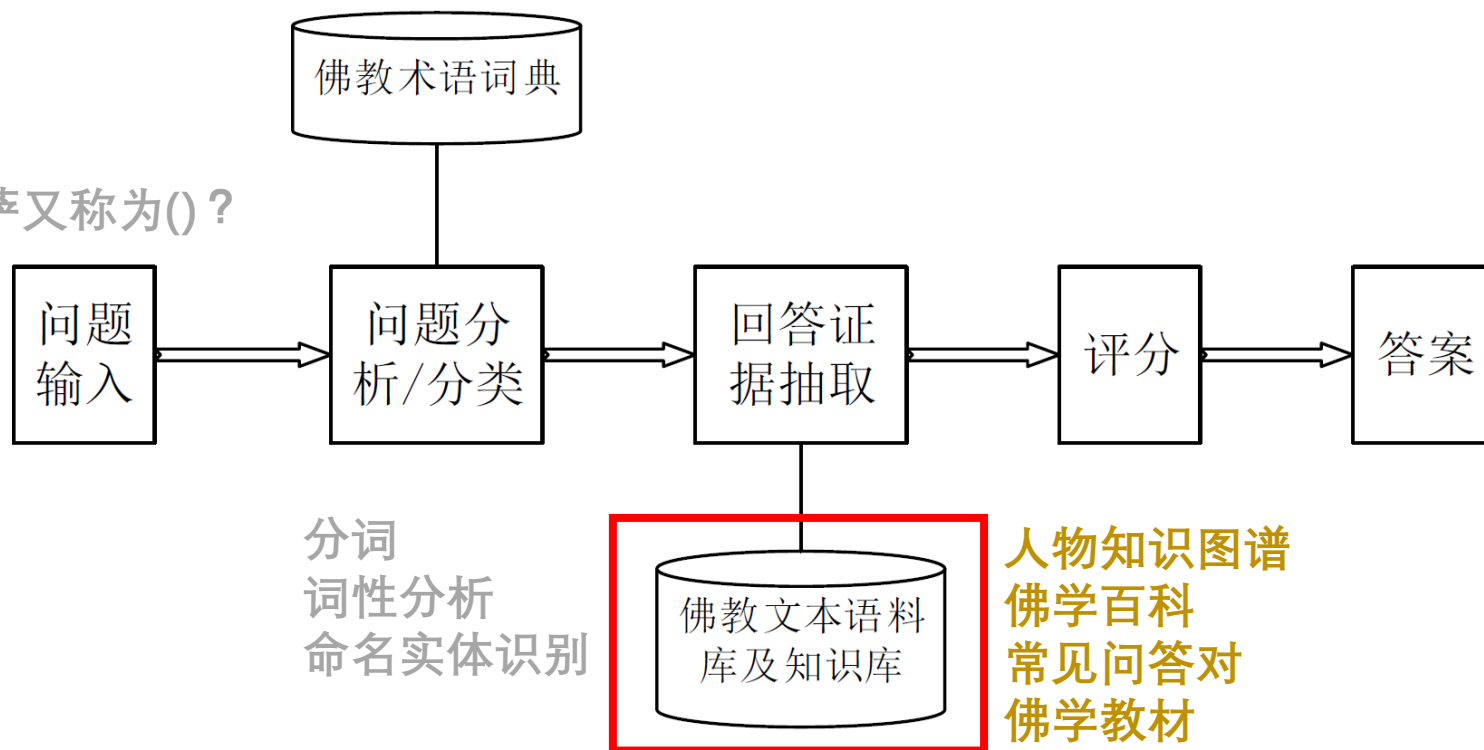


框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨

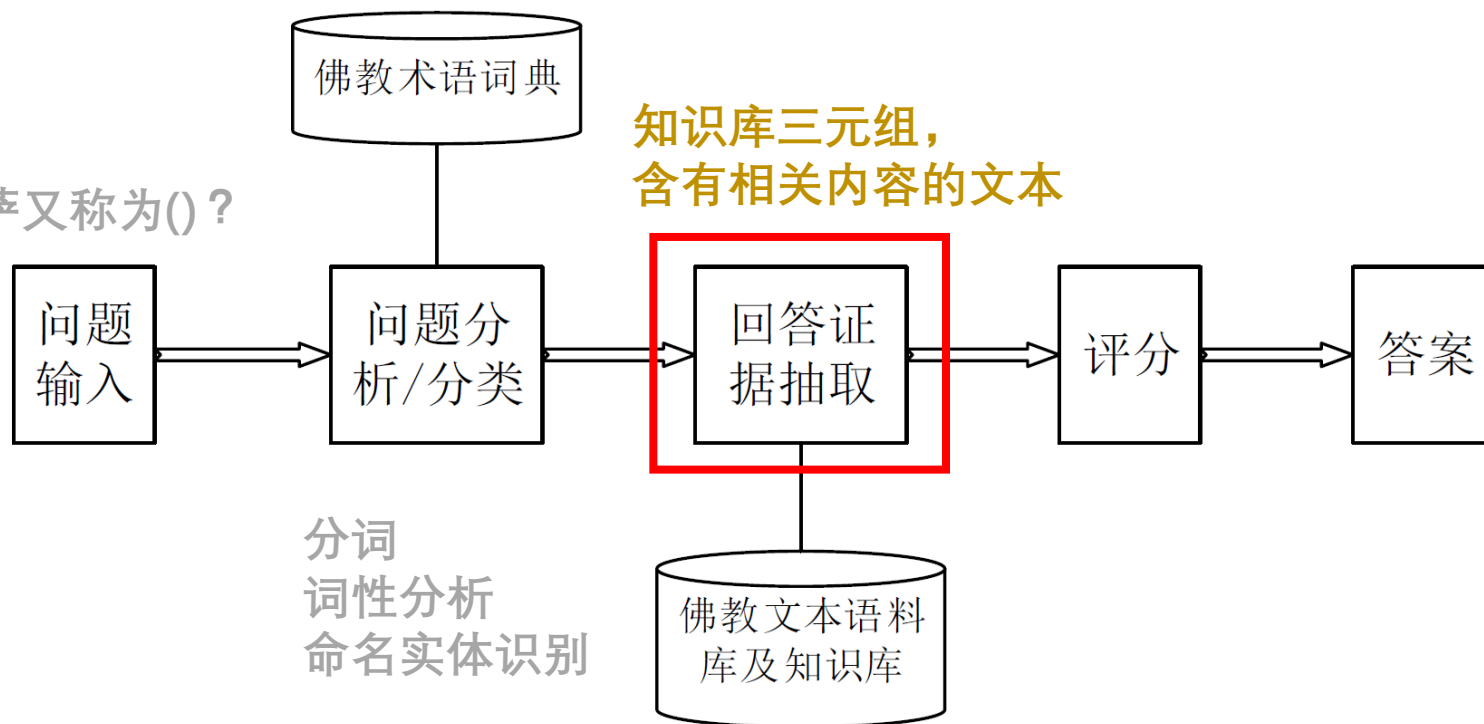


框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨



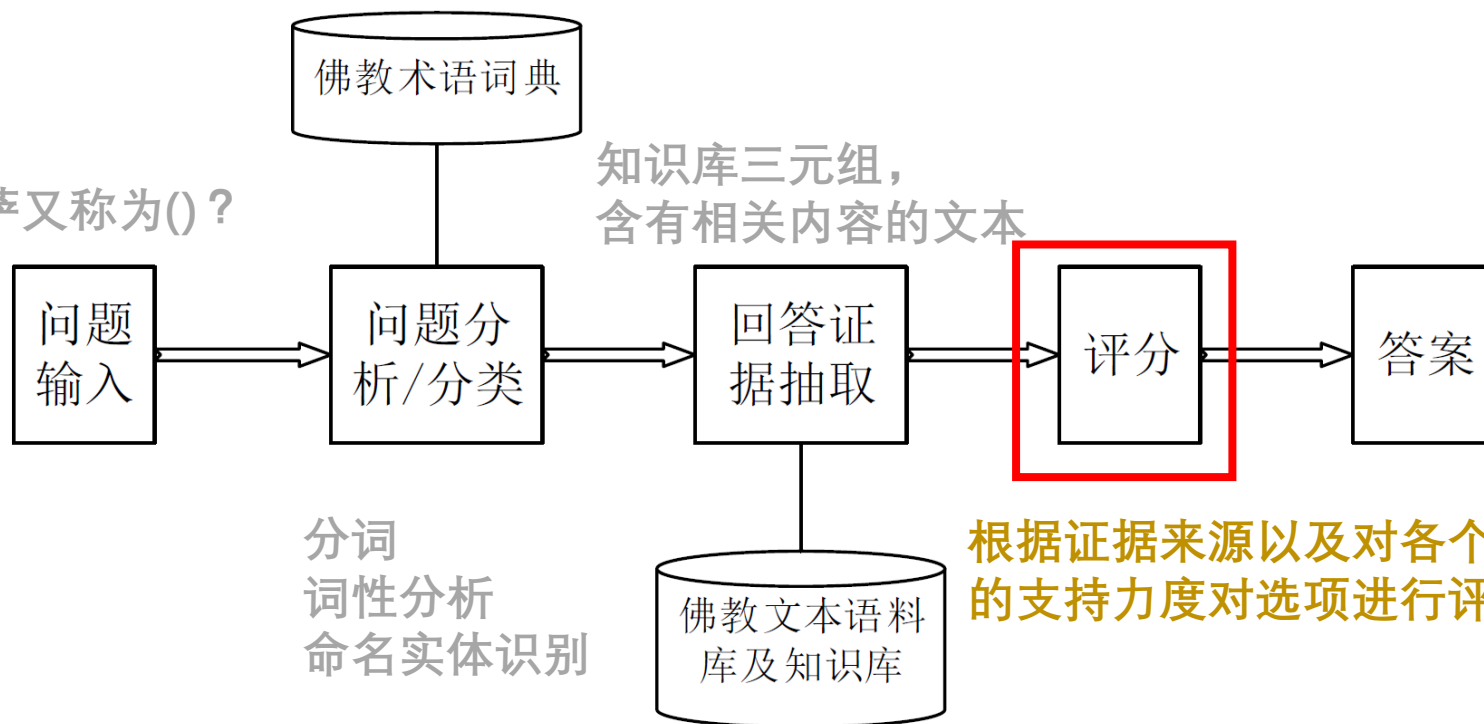
人物知识图谱、佛学百科、
常见问答对、佛学教材

框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨



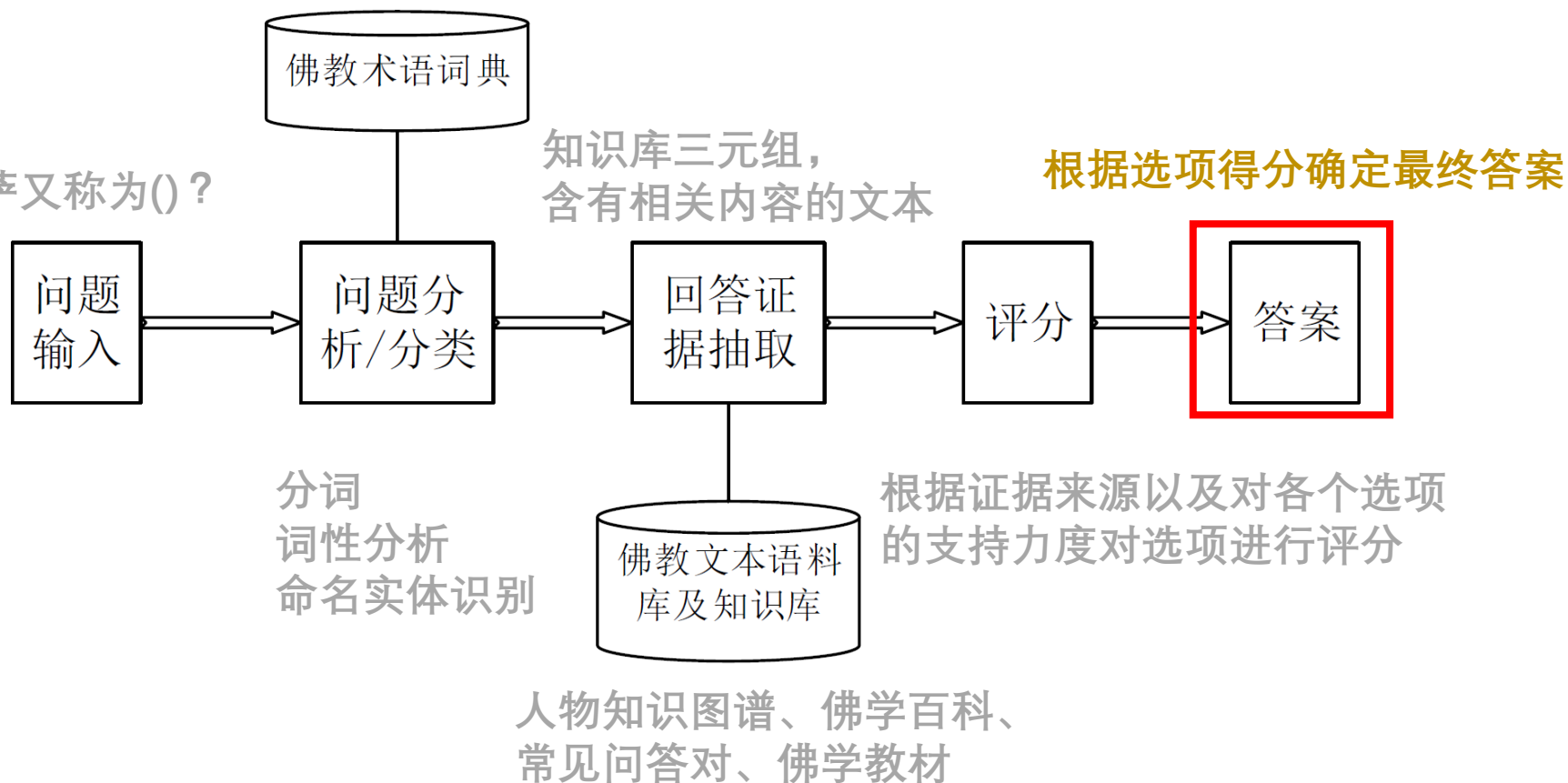
人物知识图谱、佛学百科、
常见问答对、佛学教材

框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨

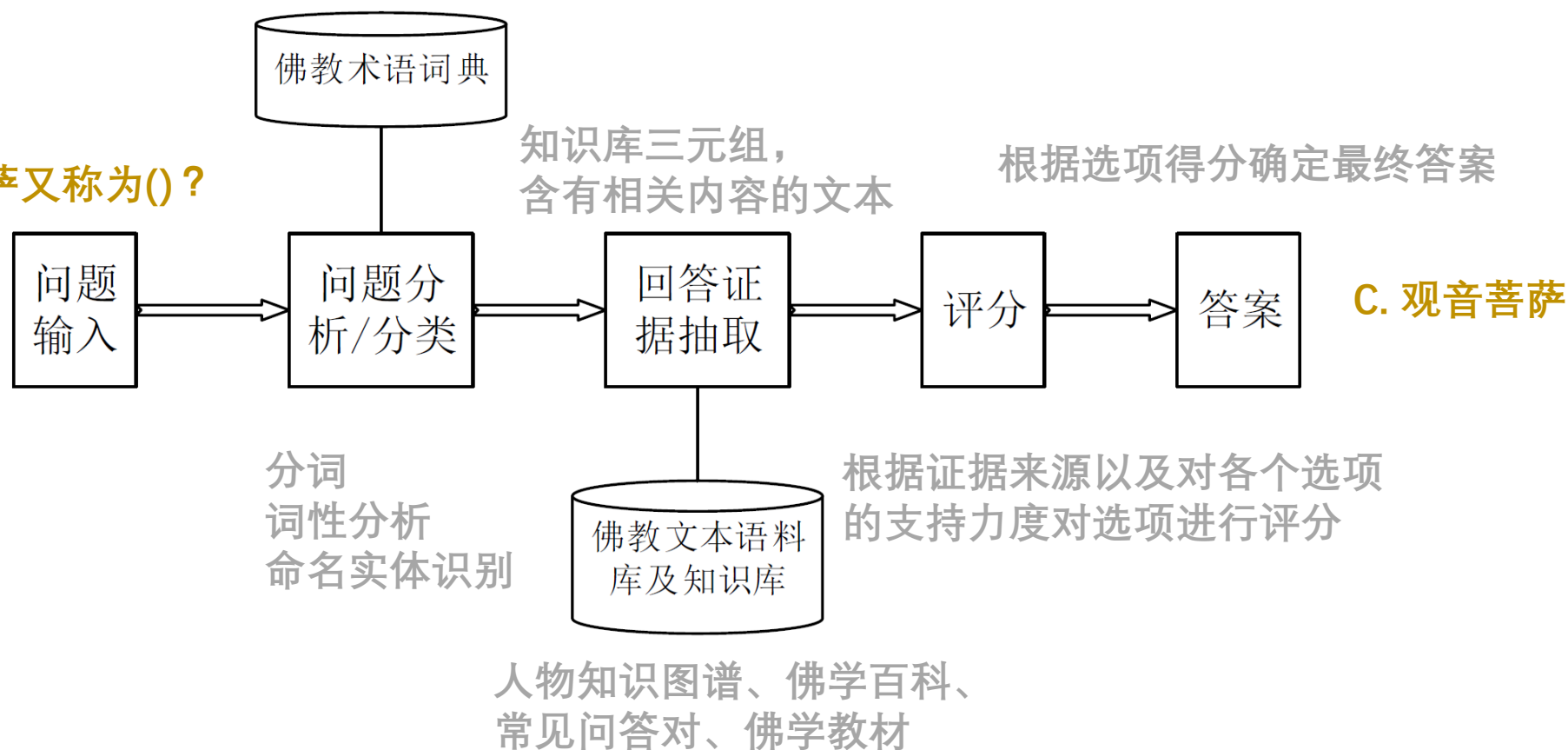


框架结构 V2.0

佛学人物词典/书籍词典

《心经》中观自在菩萨又称为() ?

- A 弥勒菩萨
- B 地藏菩萨
- C 观音菩萨
- D 文殊师利菩萨



Demo展示

佛学考试机器人

SEU-KSE

正确答案：112

问题总数：175

准确率：64%

1: 佛陀在()为五比丘说法

正确答案:鹿野苑

A.七叶窟	答案分数:0.30578512396694213
B.王舍城	答案分数:0.2892561983471074
C.鹿野苑	答案分数:0.4049586776859504

输出答案:鹿野苑

2: 佛陀告诉我们人生有()

正确答案:四苦 八苦

A.四苦	答案分数:0.23306772908366533
B.八苦	答案分数:0.23306772908366533
C.四苦 八苦	答案分数:0.5338645418326693

输出答案:四苦 八苦

总结及展望



总结

- 首个中文佛教知识图谱
- 首个针对佛学院入学考试的考试机器人
- 用技术为弘扬佛学文化尽一份力



下一步工作

佛教知识图谱

- 扩充佛教人物实体
- 改进知识图谱补全方法
- 构建寺庙知识图谱及术语知识图谱
- 构建佛学领域本体



下一步工作

佛学考试机器人

- 系统地收集各佛学院入学考试试题
- 扩充佛学相关语料，提高数据质量
- 改进问题处理及证据，答案评分方法
- 立足考试机器人，构建佛学对话系统



期待您的加入！