



上海圖書館
SHANGHAI LIBRARY
上海科技情報研究所
INSTITUTE OF SCIENTIFIC & TECHNICAL INFORMATION OF SHANGHAI



中國人民大學
RENMIN UNIVERSITY OF CHINA

The Ontology enhanced MLLM for KO&KR of multi-modal cultural memory resources

Xia Cuijuan

2024/10/20 @ NKOS2024

Outline

1. The existing methods and the GenAI-driven methods of KO&KR for multi-modal resources
2. GuQin Subtractive Notation and the multi-modal resources
3. The basic pipeline of Ontology and GenAI driven implementation for KO&KR of GuQin Subtractive Notation resources
4. The Ontology enhanced MLLM for KO&KR of GuQin Subtractive Notation resources



上海圖書館
SHANGHAI LIBRARY
上海科技情報研究所
INSTITUTE OF SCIENTIFIC & TECHNICAL INFORMATION OF SHANGHAI



中國人民大學
RENMIN UNIVERSITY OF CHINA

01

The existing methods and the GenAI-driven methods of KO&KR for multi-modal resources



01 The existing methods of KO&KR

KO(Knowledge Organization)

▲ KO 1.0

Classification
Subject Heading
Thesaurus

▲ KO 2.0

Metadata Schema
DCAP;DCAM

▲ KO 3.0

Ontology Design
CIDOC-CRM、
BIBFRAME、EDM、
SchemaBIBExt

▲ KO 4.0

Model Training for
vertical domains
LLM&MLM&MLLM

KR(Knowledge Representation)

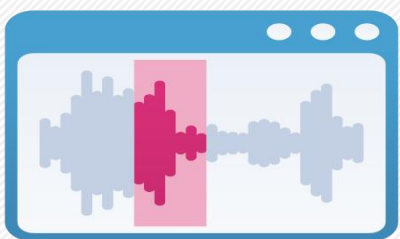
Human

Machine

The Multimodality of Cultural Memory Resources and Existing KR Method



图像 (tiff)



音频 (wav)



视频 (avi)

couldn't be caculated without KR

Textual

為一書以復朱子之舊而更以黃氏之言附益焉
世之通人以吾言為然否
左__衛成公殺叔武一段鮮
衛侯先期入
杜注不信叔武備校叔武之死成公意也前驅
逆知公意故一見而遽殺之下文元咍之訟正訟
叔武之無罪見殺也征南註此句最吃__先期
八者間道疾入不使人知
__子先長群守門以為使也與之乘而入
杜註長群衛大夫__子患公之欲速故先入欲安
喻國人備校__子知公之欲與叔武為難也故
__同隊而先入然已在公後長群不知公已間道
疾入故猶以__子為公使左氏下此三語欲見成
公之入國人始猶莫知
□□
公子過犬華仲前驅
杜註衛侯遂驅掩__子未備二子衛大夫備按
公知甯子之即在後惟慮其至則挽已
叔武將沐聞公至喜捉髮走出
備校間道疾入其至出於不意故倉卒出迎
前驅射而殺之公知其無罪也枕之股而哭之__犬
十六

text annotation (TEI+XML)

semantic

```

xmlns:cdoc="http://www.cidoc-crm.org/cidoc-crm#"
xmlns:eac-cpf="
http://archivi.ibr.regione.emilia-romagna.it/ontology/eac-cpf/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:lgdo="http://linkedgedata.org/ontology/capital" >
<rdf:Description rdf:about="
http://data.library.sh.cn/gj/resource/instance/gh7aewac687yvb8q">
<bibframe:label>抄本</bibframe:label>
<dc:type rdf:resource="
http://data.library.sh.cn/gj/vocab/sourceType/chao-ben"/>
<dc:type>善本</dc:type>
<:work rdf:resource="
http://data.library.sh.cn/entity/organization/hp0hiljxwsvdne0u"/>
<dc:title xml:lang="zh-pny">Bai Tian Chang He Ji Yi Juan</dc:title>
<pmc:classification>集 總集 地方藝文</pmc:classification>
<bibframe:extent>1</bibframe:extent>
<bibframe:edition rdf:resource="
http://data.library.sh.cn/vocab/edition/chao-ben"/>
<bibframe:contribution rdf:resource="
http://data.library.sh.cn/gj/resource/contribution/p7cept5lazzkktul"/>
<dc:subject>集 總集 地方藝文</dc:subject>
<bibframe:creator rdf:resource="
http://data.library.sh.cn/entity/person/sqj5xv2ci3kc0agn"/>
<rdf:type rdf:resource="http://pmb.library.sh.cn/ontology/Instance"/>
<dc:creator>劉南垞</dc:creator>
<bibframe:classification rdf:resource="
http://data.library.sh.cn/gj/classification/ozxoxx5uiyxqgw2"/>
<dc:title xml:lang="cht">白田倡和集一卷</dc:title>
<sh:source rdf:resource="
http://data.library.sh.cn/gj/resources/collection/9a27ihnhufwn952u"/>
<dc:title xml:lang="chs">白田倡和集一卷</dc:title>
<bibframe:instanceOf rdf:resource="
http://data.library.sh.cn/gj/resource/work/wk2dmk0h6zlk0pid"/>
<sh:description>散葉2紙</sh:description>
<bibframe:dimensions>25.3x16cm</bibframe:dimensions>
</rdf:Description>
</rdf:RDF>
    
```

ontology and semantic data (RDF/KG)

keyword match retrieval

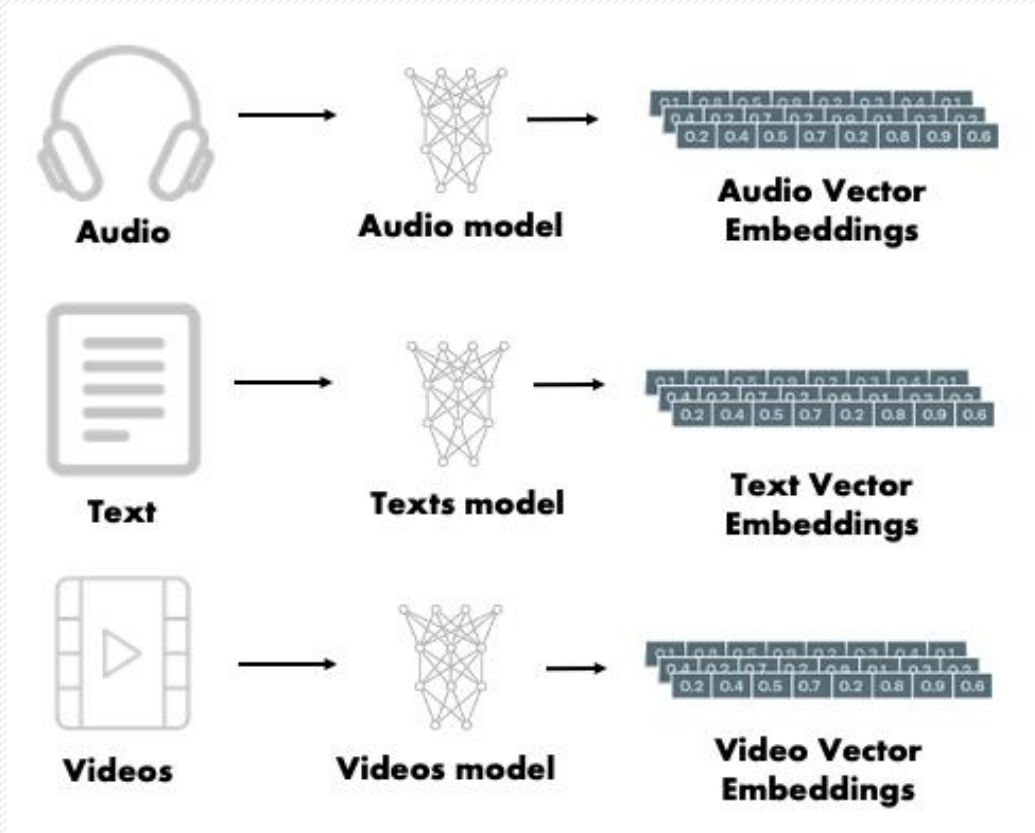
entigy/concept match retrieval

structured

文献类型	文种	收藏单位名称	馆藏索书号	全文索书号
敦煌遗书	汉文	上海圖書館	線善812441	812441
敦煌遗书	汉文	上海圖書館	線善812494	812494
汉文古籍	汉文	上海圖書館	線善829929-31	829929-31
汉文古籍	汉文	上海圖書館	線善830012-17	830012-17
汉文古籍	汉文	上海圖書館	線善753995-4002	753995-4002
汉文古籍	汉文	上海圖書館	線善758489	758489
汉文古籍	汉文	上海圖書館	線善754003-06	754003-06
汉文古籍	汉文	上海圖書館	線善829046-53	829046-53
汉文古籍	汉文	上海圖書館	線善829822-30	829822-30
汉文古籍	汉文	上海圖書館	線善828904-08	828904-08
汉文古籍	汉文	上海圖書館	線善754011-15	754011-15
汉文古籍	汉文	上海圖書館	863326-35	863326-35
汉文古籍	汉文	上海圖書館	線善754016-21	754016-21
汉文古籍	汉文	上海圖書館	線善795539-48	795539-48
汉文古籍	汉文	上海圖書館	線善795549-51	795549-51

medadata schema and structured data (csv)

GenAI-driven methods of KO&KR for multi-modal resources



Similarity Retrieval & Cross Modal Retrieval & Natural Language Retrieval



上海圖書館
SHANGHAI LIBRARY
上海科技情報研究所
INSTITUTE OF SCIENTIFIC & TECHNICAL INFORMATION OF SHANGHAI

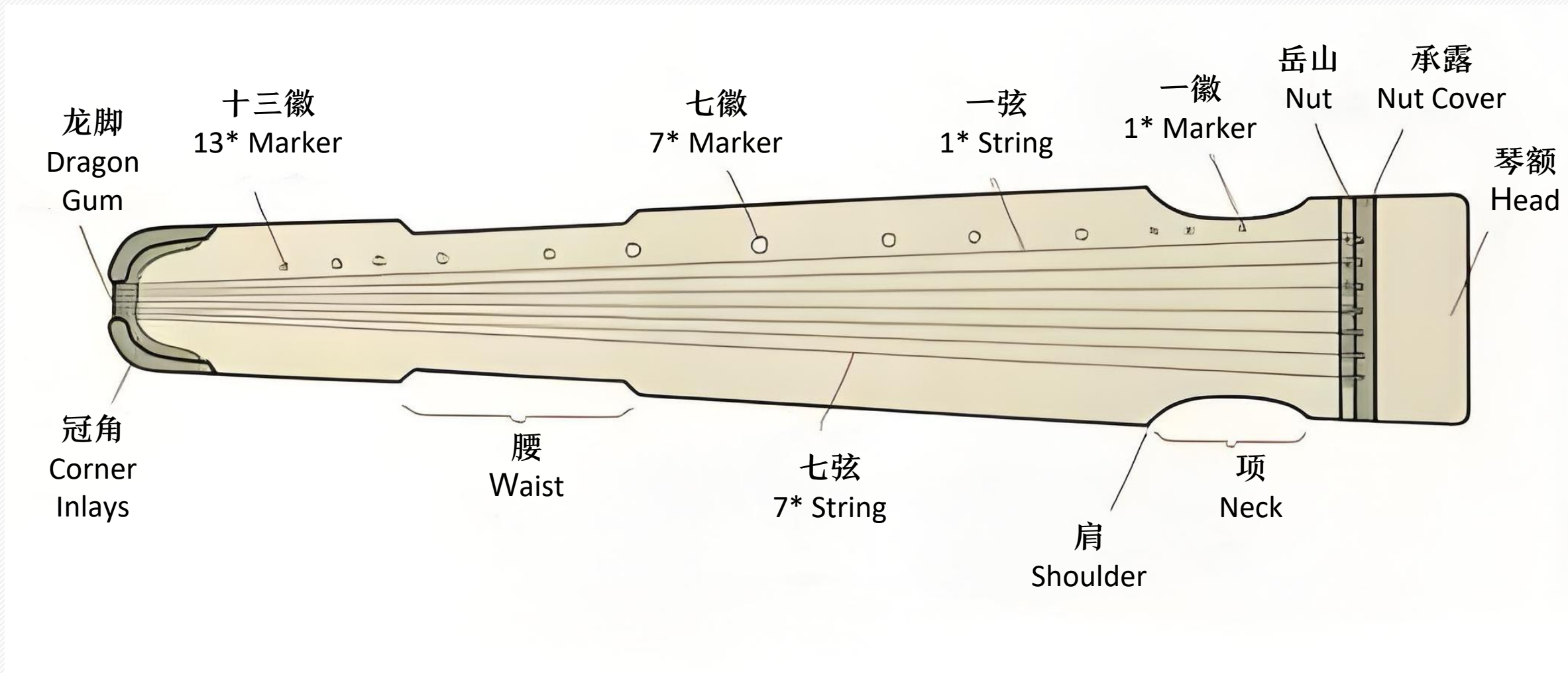


中國人民大學
RENMIN UNIVERSITY OF CHINA

02

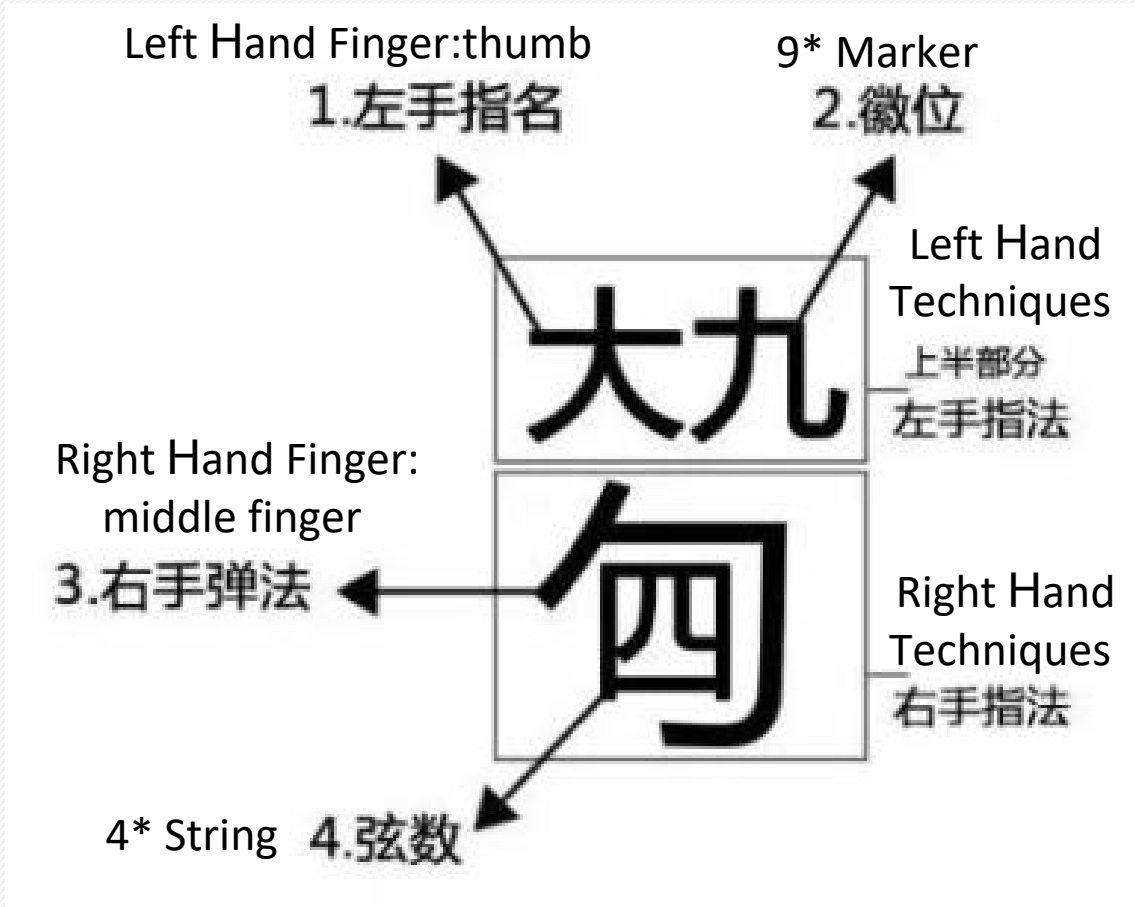
GuQin Subtractive Notation and the multi-modal resources





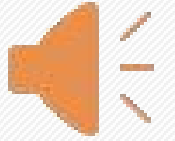
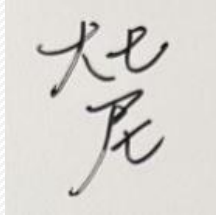


GuQin Subtractive Notation and the multi-modal resources



大七
擘七

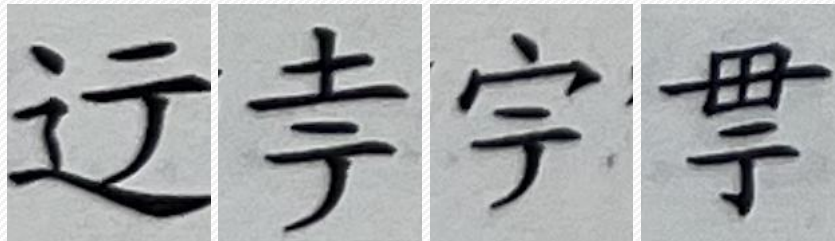
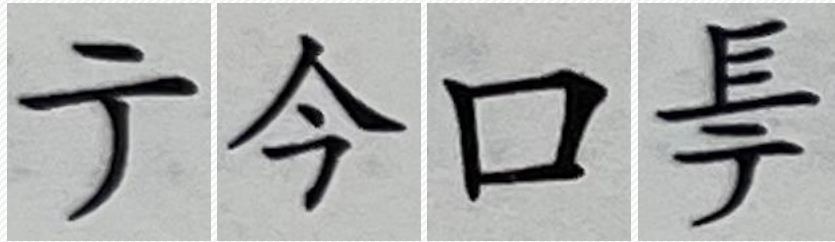
大七
尾





Ambiguity, ambiguity and uncertainty of GuQin Subtractive Notation

multiple characters of one notation



different playing styles of one notation



吟	【解】左指按某徽，右指弹后，左指缓缓摇动，叫吟。用甲按叫甲吟，用肉按叫肉吟。其法有二：①当右指弹后，左指随声作吟。②得声后稍停再吟，叫见声吟或声后吟，即近代谱中的落指吟或缓吟。	左手指法
长吟	【解】仅上下滑动的时间较一般吟长，叫长吟。	左手指法
游吟	【解】左大指（中名指亦可）按某弦某徽，当右指弹后，在所按处上下往来滑动，先缓后急，叫游吟。	左手指法
走吟	【解】左指得音后，滑上或滑下就吟，叫走吟。	左手指法
实吟	【解】左名指按弦得音后。微微滑动，叫实吟。	左手指法
吟上	【解】吟后，随即滑上，叫吟上。	左手指法
吟下	【解】左指吟后，随即滑下，叫吟下。	左手指法
注吟	【解】左指得音后，立即注下至某徽位作吟，叫注吟。	左手指法



Scenarios of KO&KR for GuQin Subtractive Notation resources

Scenario 1: Notation recognition

- **Notation character recognition:**
identifying Notation names
- **Music Notation recognition:**
context recognition to identify Notation characters

Scenario 2: Notation interpretation

Explain the playing techniques and styles

Scenario 3: Notation transformation and creation

Scenario 4: Cross modal retrieval

Notation & Music Search cross-modalities

Scenario 5: Cross modal generation

- **Teaching and training:**
Recognize and Interpret the Music Notation, Distinguish the sounds one played, Demonstrate how to play it correctly.
- **Music creation:**
Generation of new music (music notation, audio, video)



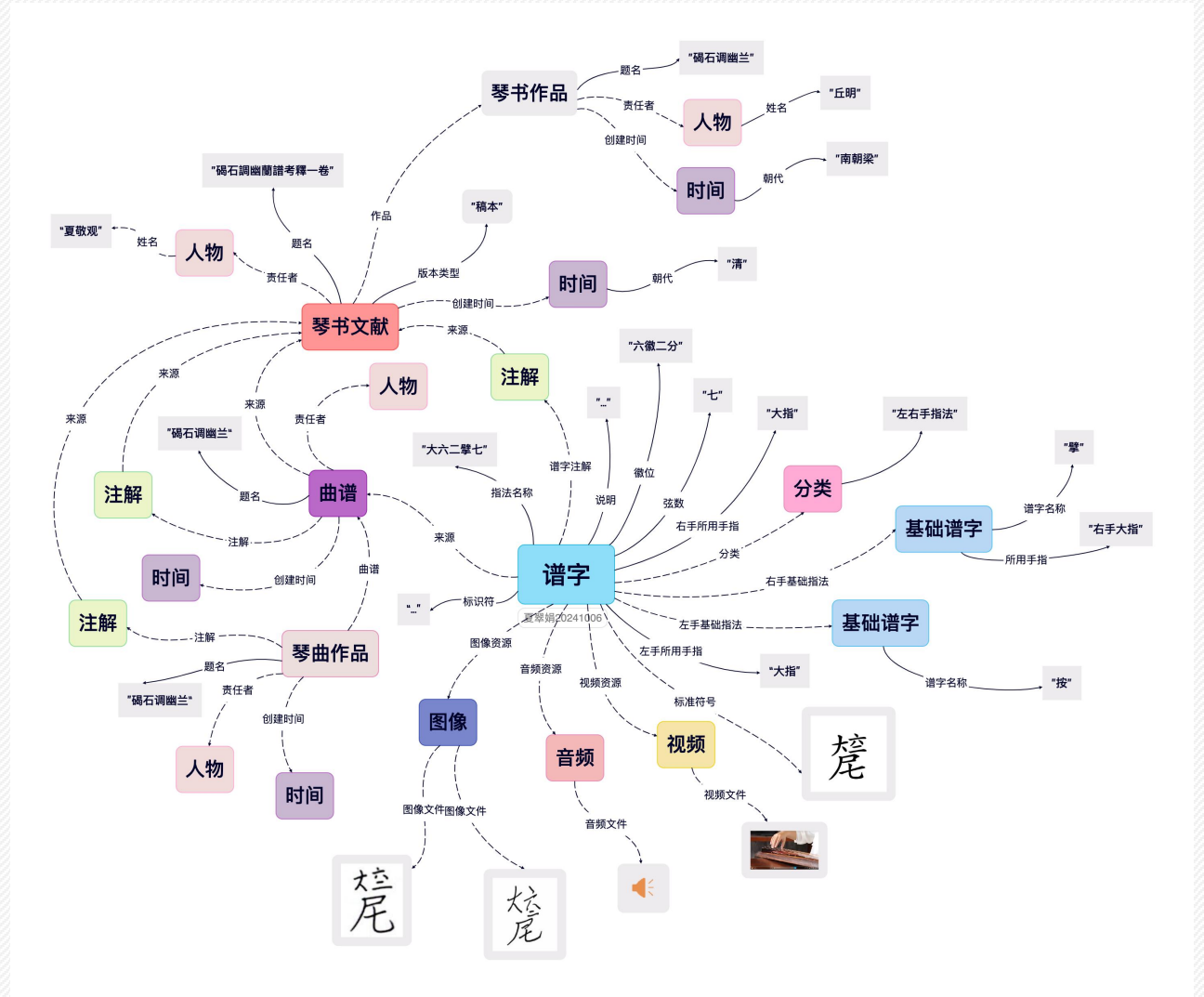
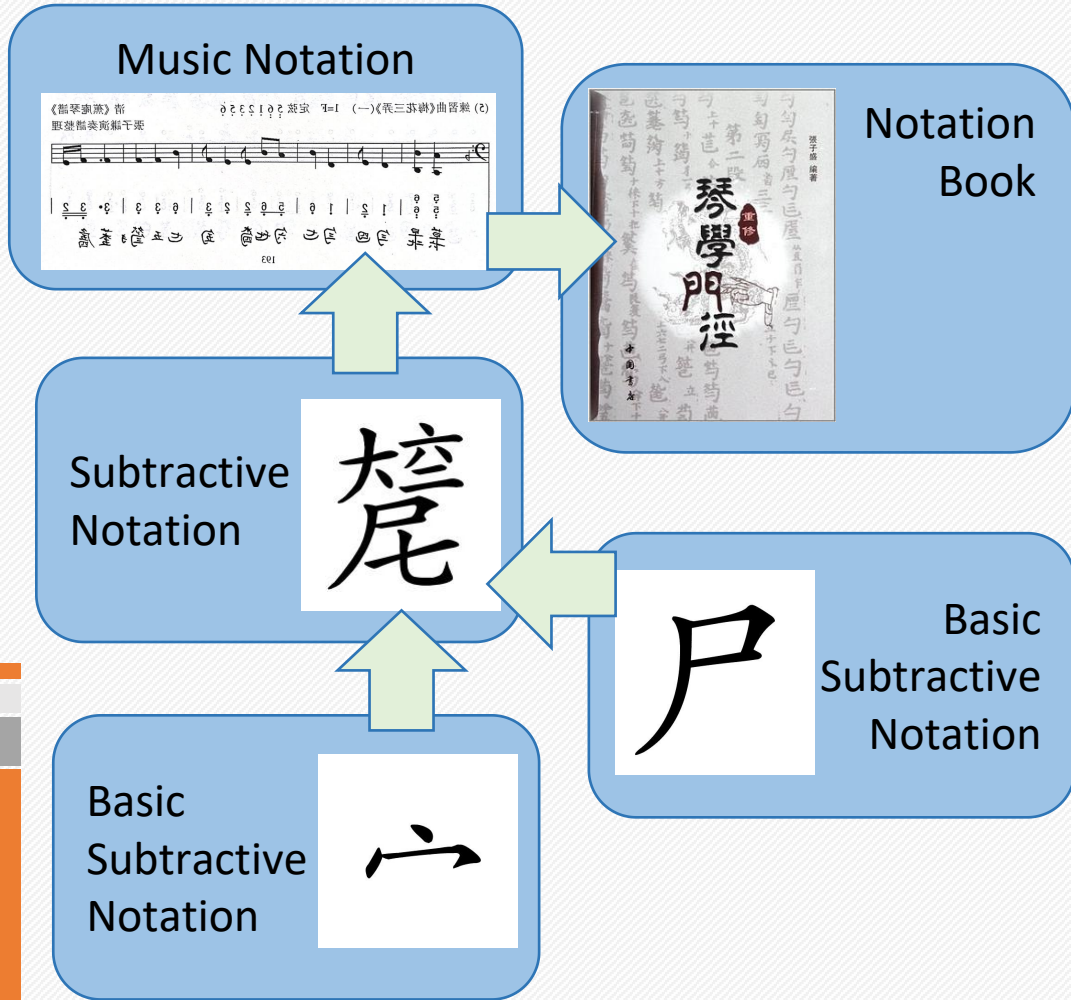
03

The basic pipeline of Ontology and GenAI driven
implimentation for KO&KR for GuQin Subtractive
Notation resources



03 Ontology-driven KO

Ontology Model



03 Ontology-driven KO

Ontology Vocabulary

类 Class	标签 Label	父类 SubClassOf
文化记忆资源	gq:CulturalMemoryResource	rdf:Resource
谱字	gq:SubtractiveNotation	gq:CulturalMemoryResource
图像	gq:Image	gq:CulturalMemoryResource
音频	gq:Audio	gq:CulturalMemoryResource
视频	gq:Video	gq:CulturalMemoryResource
文献作品	gq:Work	gq:CulturalMemoryResource
琴书文献	gq:Instance	gq:CulturalMemoryResource
曲谱	gq:MusicNotation	gq:CulturalMemoryResource
乐曲作品	gq:Music	gq:CulturalMemoryResource
注解	gq:Annotation	gq:CulturalMemoryResource
人物	gq:Person	foaf:Person
时间	gq:Temporal	time:Temporal
基础谱字	gq:BasicNotation	gq:SubtractiveNotation
图像资源	gq:Image	gq:CulturalMemoryResource;shema:ImageObject
音频资源	gq:Audio	gq:CulturalMemoryResource;schema:AudioObject
视频资源	gq:Video	gq:CulturalMemoryResource;schema:VideoObject

属性Property	标签Label	属性类型PropertyType	域Domain	范围Range
谱字名称	gq:notationName	Data Property	gq:SubtractiveNotation	rdfs:Literal
说明	dc:description	Data Property	gq:CulturalMemoryResource	rdfs:Literal
徽数	gq:marker	Data Property	gq:SubtractiveNotation	rdfs:Literal
弦数	gq:string	Data Property	gq:SubtractiveNotation	rdfs:Literal
所用手指	gq:finger	Data Property	gq:SubtractiveNotation	rdfs:Literal
分类	gq:category	Data Property	gq:SubtractiveNotation	rdfs:Literal
右手基础指法	gq:rightHandTech	Object Property	gq:SubtractiveNotation	gq:SubtractiveNotation
左手基础指法	gq:leftHandTech	Object Property	gq:SubtractiveNotation	gq:SubtractiveNotation
标准符号	gq:standardChar	Object Property	gq:SubtractiveNotation	gq:Image
视频资源	gq:videoResource	Object Property	gq:SubtractiveNotation	gq:Video
音频资源	gq:audioResource	Object Property	gq:SubtractiveNotation	gq:Audio
图像资源	gq:imageResource	Object Property	gq:SubtractiveNotation	gq:Image
创建时间	dct:created	Data Property	gq:CulturalMemoryResource	gq:Temporal
责任者	dc:creator	Data Property	gq:CulturalMemoryResource	rdfs:Literal
责任者	dct:creator	Object Property	gq:CulturalMemoryResource	gq:Person
题名	dc:title	Data Property	gq:CulturalMemoryResource	rdfs:Literal
来源	dct:source	Object Property	gq:CulturalMemoryResource	gq:Instance
注解	gq:annotation	Object Property	gq:CulturalMemoryResource	gq:Annotation
曲谱	gq:hasMusicNotation	Object Property	gq:Music	gq:MusicNotation
姓名	foaf:name	Data Property	foaf:Person	rdfs:Literal
别名	gq:AbbreviateName	Data Property	foaf:Person	rdfs:Literal
作品	gq:isInstanceOf	Object Property	gq:Instance	gq:Work

03 Ontology-driven KR

Application Profile

属性 Property	标签 Label	域 Domain	范围 Range	用于描述 UsedWith	预期取值 ExpectedValue	必备性 Mandatory	可重复性 Repeatable
谱字名称	gq:notationName	gq:SubtractiveNotation	rdfs:Literal	gq:SubtractiveNotation	rdfs:Literal	Y	N
说明	dc:description	gq:CulturalMemoryResource	rdfs:Literal	gq:CulturalMemoryResource	rdfs:Literal	N	N
徽数	gq:marker	gq:SubtractiveNotation	rdfs:Literal	gq:SubtractiveNotation	rdfs:Literal	N	Y
弦数	gq:string	gq:SubtractiveNotation	rdfs:Literal	gq:SubtractiveNotation	rdfs:Literal	N	Y
所用手指	gq:finger	gq:BasicNotation	rdfs:Literal	gq:SubtractiveNotation	rdfs:Literal	N	Y
分类	gq:category	gq:SubtractiveNotation	rdfs:Literal	gq:SubtractiveNotation	rdfs:Literal	Y	Y
右手基础指法	gq:rightHandTech	gq:SubtractiveNotation	gq:SubtractiveNotation	gq:SubtractiveNotation	gq:SubtractiveNotation	N	Y
左手基础指法	gq:leftHandTech	gq:SubtractiveNotation	gq:SubtractiveNotation	gq:SubtractiveNotation	gq:SubtractiveNotation	N	Y
标准符号	gq:standardChar	gq:SubtractiveNotation	gq:Image	gq:SubtractiveNotation	gq:Image	Y	N
视频资源	gq:videoResource	gq:SubtractiveNotation	gq:Video	gq:SubtractiveNotation	gq:Video	N	Y
音频资源	gq:audioResource	gq:SubtractiveNotation	gq:Audio	gq:SubtractiveNotation	gq:Audio	N	Y
图像资源	gq:imageResource	gq:SubtractiveNotation	gq:Image	gq:SubtractiveNotation	gq:Image	N	Y
创建时间	dct:created	gq:CulturalMemoryResource	rdfs:Literal	gq:CulturalMemoryResource	gq:Temporal	N	N
责任者	dc:creator	gq:CulturalMemoryResource	rdfs:Literal	gq:CulturalMemoryResource	rdfs:Literal	Y	Y
责任者	dct:creator	gq:CulturalMemoryResource	gq:Person	gq:CulturalMemoryResource	gq:Person	Y	Y
题名	dc:title	gq:CulturalMemoryResource	rdfs:Literal	gq:CulturalMemoryResource	rdfs:Literal	Y	Y
来源	dct:source	gq:CulturalMemoryResource	gq:Instance	gq:CulturalMemoryResource	gq:Instance	N	Y
注解	gq:annotation	gq:CulturalMemoryResource	gq:Annotation	gq:CulturalMemoryResource	gq:Annotation	N	Y
曲谱	gq:hasMusicNotation	gq:Music	gq:MusicNotation	gq:Music	gq:MusicNotation	Y	Y
姓名	foaf:name	foaf:Person	rdfs:Literal	foaf:Person	rdfs:Literal	Y	N
别名	gq:AbbrivateName	foaf:Person	rdfs:Literal	foaf:Person	rdfs:Literal	Y	Y
作品	gq:InstanceOf	gq:Instance	gq:Work	gq:Instance	gq:Work	Y	N



03 Ontology-driven KR

Encoding as Triples

<sn1> a gq:SubtractiveNotation;

```

dc:identifier "sn1";
gq:notationName "大八半擘四";
gq:category "左右手谱字";
dc:description "用左手大指按住八徽五分的同时;用右手大指由内向外擘四弦。";
gq:annotation [a gq:Annotation; dcsource "ChatGPT 4o".>;
gq:standardChar [a gq:Image; dc:source "上海音乐出版社楷体字符集".]
gq:rightHandBasicNotation <bsn1>;
gq:string "四";
gq:leftFinger "大指";
gq:leftHandBasicNotation <bsn2>;
gq:marker "八徽五分";
gq:rightFinger "大指".

```

<bn1> a gq:BasicNotation;

```

dc:identifier "bsn1";
gq:notationName "擘";
dc:description "右手大指由内向外擘弦。";
gq:finger "大指";
gq:category "右手谱字","基础指法".

```

<bn2> a gq:BasicNotation;

```

dc:identifier "bsn2";
gq:notationName "按";
dc:description "左手按弦得声，为实音。";
gq:category "左手谱字","基础指法".

```

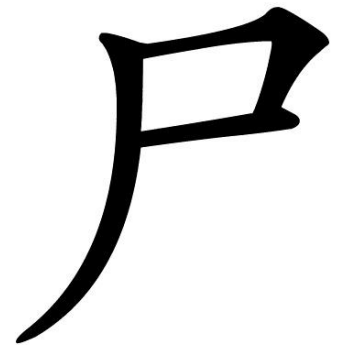
<sn1> a gq:SubtractiveNotation



<bn1> a gq:BasicNotation



<bn2> a gq:BasicNotation



03 GenAI—driven

Prompt Engineering

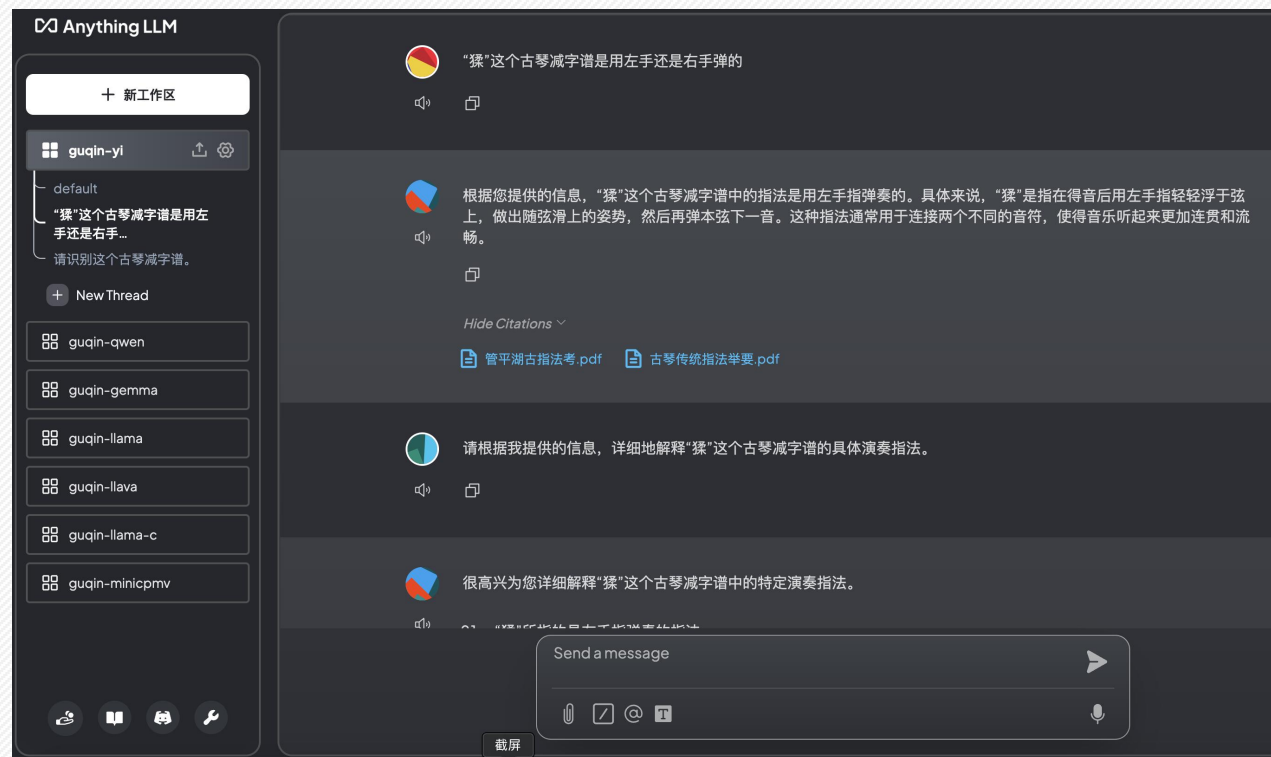
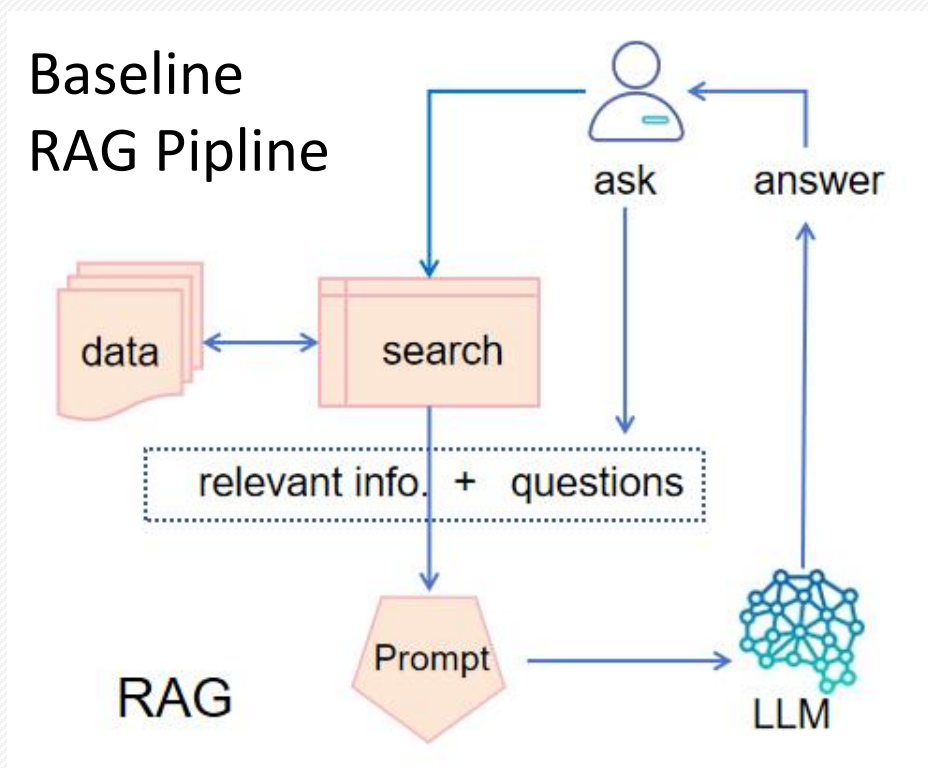
Test 1: Does the existing MLLMs have the basic knowledge of the vertical domain of Guqin Subtractive Notation?

Test 2: Can the existing MLLMs recognize the Guqin Subtractive Notation images and make correct explanations?

Test 3: Can the existing MLLMs achieve cross modal generation of Guqin Subtractive Notation?



Conclusion: Without professional knowledge of Guqin Subtractive Notation, it is impossible for the MLLMs to recognize the notation images; or to achieve multimodal generation; By prompt engineering, the MLLMs can be optimized. But Only ChatGPT 4o can provide correct interpretation of images, even has memory for the user personally.



Conclusion: The baseline RAG can supplement professional knowledge about Guqin Substractive Notation for large models. And it performs well in discriminative tasks such as binary classification. However, for generative and interpretive tasks, it is difficult to utilize the local knowledge in the provided documents.

Xia Cuijuan. Phased achievements of doctoral dissertation research of School of Information Resources Management; Renmin University of China



04

The Ontology enhanced MLLM for for KO&KR of
GuQin Subtractive Notation resources

Finetune & Demo



04 Finetune (Embedding Model)

Low-Rank Adaptation

Fine tuning Experiment 1:

Model size: 2B; with 5% adjustable parameters

Hardware: Apple Pro M3 48G CPU

Dataset: 96 GuQin Subtractive Notation with vision-text comparison data (500 images)

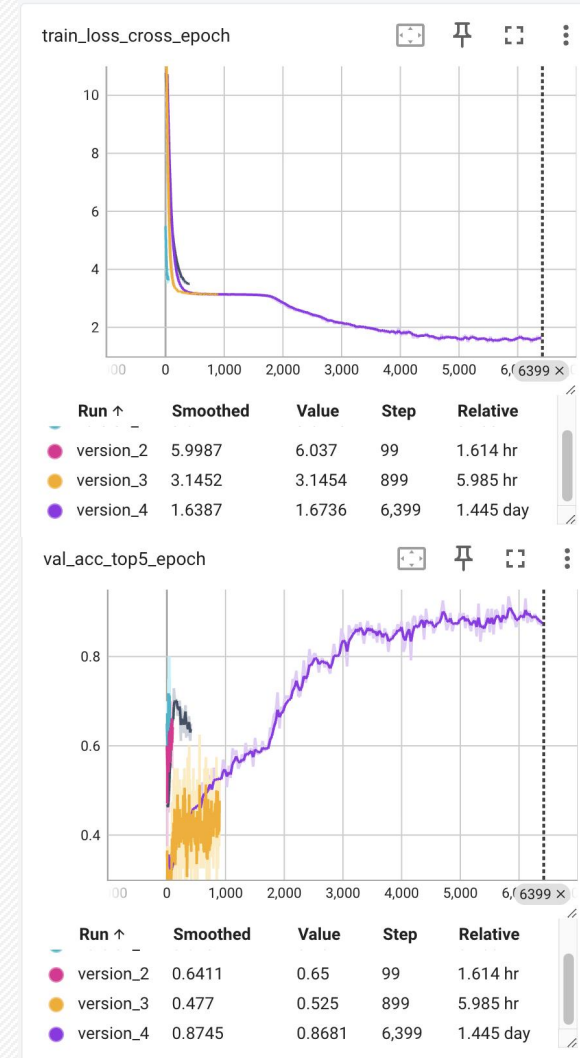
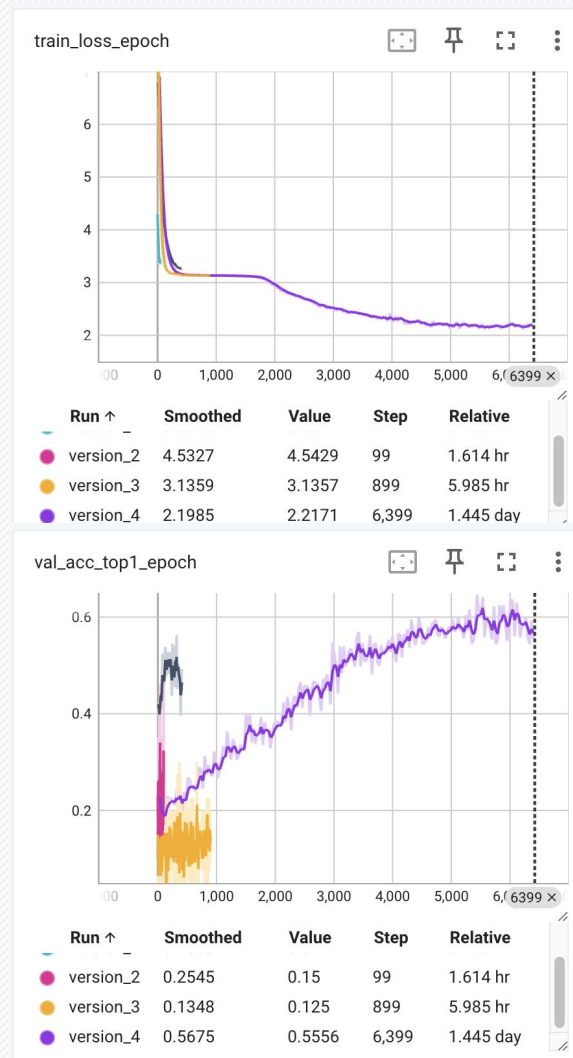
train: val 8:2

Fine tuning time: 200 epochs/36 hours

Hyperparameter setting: lr 5e-6

Fine tuning results: epoch loss : 2.2171 ;cross epoch

loss: 1.6736 ; acc top 1: 0.5556 ; acc top 5: 0.8681



04 Finetune (Embedding Model)

Experiment 1 Demo

跨模态的向量相似性检索
Vector Similarity Search Across Audio, Video, Text, Image based on MLLM

by Cuijuan(Jada) Xia2024-10-19 15:27:38

Choose a model

Input a text query

Input an image query

将图像拖放到此处
- 或 -
点击上传

Input an audio query

将音频拖放到此处
- 或 -
点击上传

Clear Submit

Text

Audio

Standard Image

Video

Flag

04 Finetune (Embedding Model)

Low-Rank Adaptation

Fine tuning experiment 2:

Model size: 2B; with 5% adjustable parameters

Hardware: NVIDIA 4090D 24G

Dataset: 96 GuQin Subtractive Notation with vision-text comparison data (4000 images)

train: val 8:2

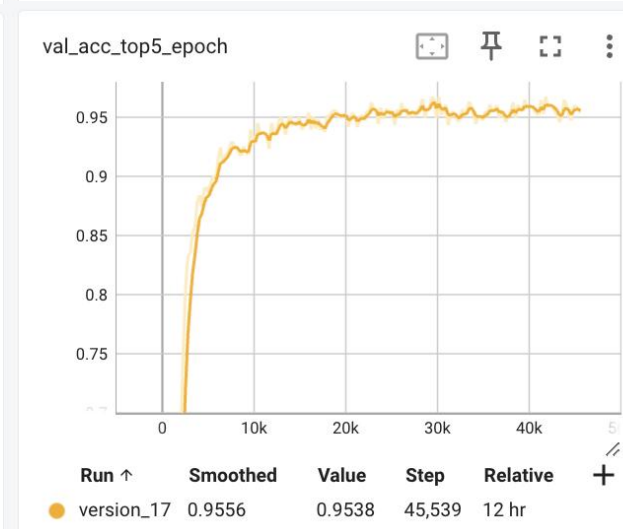
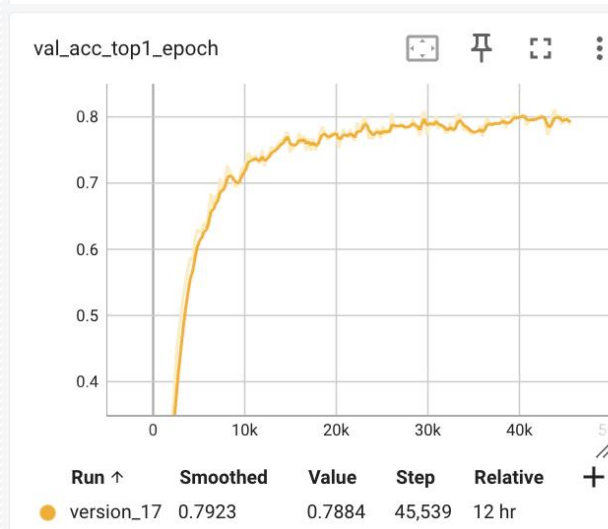
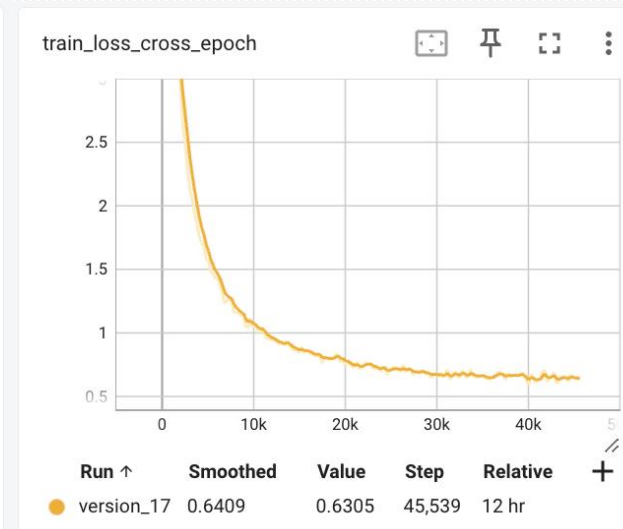
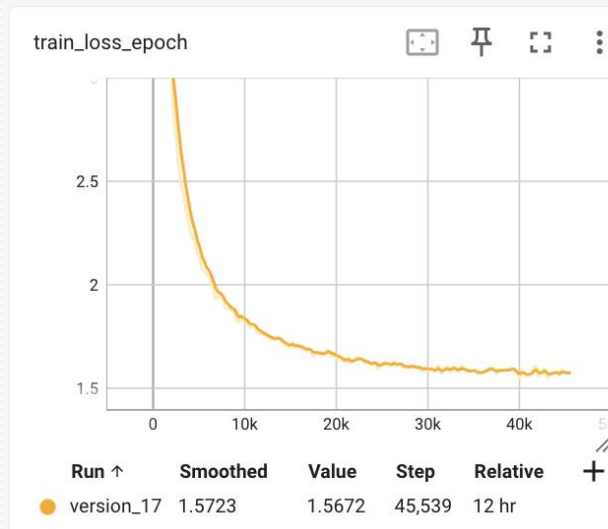
Fine tuning time: 180 epochs/12 hours

Hyperparameter setting: lr 5e-6

Fine tuning results:

epoch loss : 1.5672 ↓ cross epoch loss: 0.6305 ↓

acc top 1: 0.7923 ↑ acc top 5: 0.9556 ↑



04 Finetune (Embedding Model)

Experiment 2 Demo

跨模态的向量相似性检索
Vector Similarity Search Across Audio, Video, Text, Image based on MLLM

by Cuijuan(Jada) Xia2024-10-19 15:53:58

Choose a model

GuQin_IB-LoRA-180

Input a text query

Input an image query

将图像拖放到此处
-或-
点击上传

Input an audio query

将音频拖放到此处
-或-
点击上传

Clear

Submit

Text

Audio

Standard Image

Video

Flag

04 Finetune (Chat Model) with text data as instruction prompts

Fine tuning Experiment 3:

Model size: 2B; adjustable parameters

5%

Hardware: NVIDIA 4090D 24G

Dataset: 1W+GuQin Subtractive

Notation with 1W+ images and

125W+word prompts

train: val 8:2

Fine tuning time: 160 epochs/36 hours

Hyperparameter setting: lr 5e-6

```
prompts_and_classes.txt
'a {0} {1} 指法名称: 吟。指法分类: 左手。指法描述: 吟也, 《五知斋》《自远堂》皆注云: 如理自速上注, 《琴学圣》略同, 《松风阁》云: 即轻逗, 贵无形迹。今据《松风阁》用反撞而少轻, 亦贵无迹也。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 临。指法分类: 左手。指法描述: 临也, 各谱已删不用之指法, 《五知斋》列入古指法, 注云: “食指一路滚去。”《自远堂·水仙操》一曲相沿用之未改, 今仍照刊作滚弹。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 微吟。指法分类: 左手。指法描述: 微吟也, 亦旧谱已删不用之指法, 《自远堂》列入旧指法录存内。《水仙》原本有此字, 今仍之 (其余指法, 详见原谱)。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 吟。指法分类: 左手。指法描述: 吟也, 有五吟、四吟两种。五吟者, 从本位退下二分, 复上至本位, 又作两退复, 若唱弦, 则略去首一退字, 唱为“复退复退复”, 论板, 则第一复字一板, 第二退字一板, 《琴镜》谱旁注写为“于于于于”是也。四吟即如前法, 唱为“退复退复”, 论板, 则第二退字一板, 本谱旁注写为“于于于于”是也, 其余吟法以此类推。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 游吟。指法分类: 左手。指法描述: 游吟也, 得音后自本位退至下位, 复上至本位, 又退至下位, 又复上本位, 再加退二分之退复退复。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 飞吟。指法分类: 左手。指法描述: 飞吟也, 得音后飞至上位, 即复本位, 再加退复退复, 此照《徵言秘旨订》法也。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 荡吟。指法分类: 左手。指法描述: 荡吟也, 得音后退下二分, 用力复本位, 连作退复退复四声, 又作退复二声, 作复俱用力, 退不用力。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 缓吟。指法分类: 左手。指法描述: 缓吟也, 得音后过一板, 乃用吟, 缓猱同。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 双吟。指法分类: 左手。指法描述: 双吟也, 如七弦则唱: 七退复退七复退复退复。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 往来吟。指法分类: 左手。指法描述: 往来吟也, 有得音后上一位, 即就上一位作本位, 加退复退复退者, 有就本位用退复退复退者。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 定吟。指法分类: 左手。指法描述: 定吟也, 大吟、大猱后, 按弦不动, 过一板弹下声。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 细吟。指法分类: 左手。指法描述: 细吟也, 得音后急上二分, 急下二分, 急复本位。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 猱。指法分类: 左手。指法描述: 猱也, 得音后, 就本位进二分, 复本位, 再作两边, 为三进复, 皆以复为猱, 猱必在本位之上, 吟在本位之下。《琴粹》以猱再本位下, 吟在本位上下, 实传习之误, 《琴镜·羽化登仙跋》中已更正, 须知弹琴指法之妙全在吟猱, 欲扬以连下声则用猱, 欲抑以连下声则用吟, 谱中猱之种类多以板别之, 不赘录。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 唤。指法分类: 左手。指法描述: 唤也, 得音后急作分许小退复, 略如反撞。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 进复。指法分类: 左手。指法描述: 进复也。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 退复。指法分类: 左手。指法描述: 退复也。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 懒进复。指法分类: 左手。指法描述: 懒进复也, 进用一板或二板为懒。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 逗。指法分类: 左手。指法描述: 逗也, 如九徽之音, 从八徽弹下, 逢逗必速, 过重则俗, 尤忌多用。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 指起。指法分类: 左手。指法描述: 指起也, 名指按十徽, 大指按九徽, 指起得声。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 对起。指法分类: 左手。指法描述: 对起也, 右有弹而左爪起为对起, 无弹爪起为爪起。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 爪起。指法分类: 左手。指法描述: 爪起也, 大指甲将弦爪起, 得一散声。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 带起。指法分类: 左手。指法描述: 带起也, 名指甲带弦起, 得一散声。'.format(unique_token, class_token),
'a {0} {1} 指法名称: 跪。指法分类: 左手。指法描述: 跪也, 名指屈末节, 以甲按弦, 连跪两弦, 兼用节。'.format(unique_token, class_token),
```


04 Finetune (Chat Model) with text data as instruction prompts

Fine tuning Experiment 3:

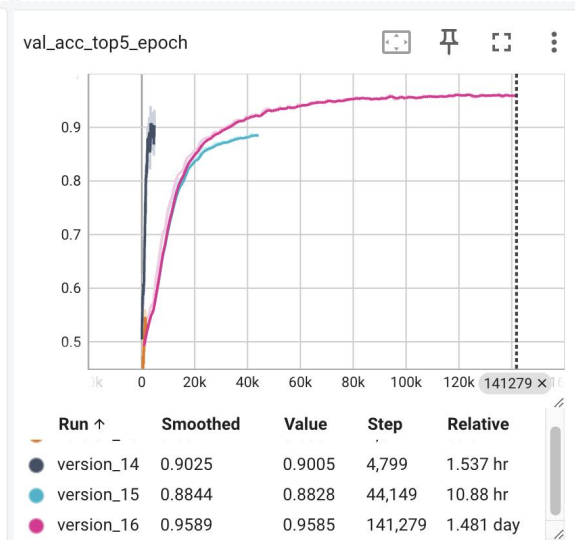
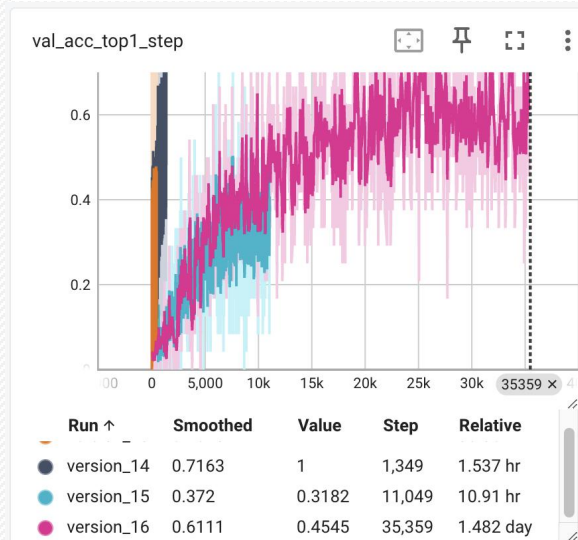
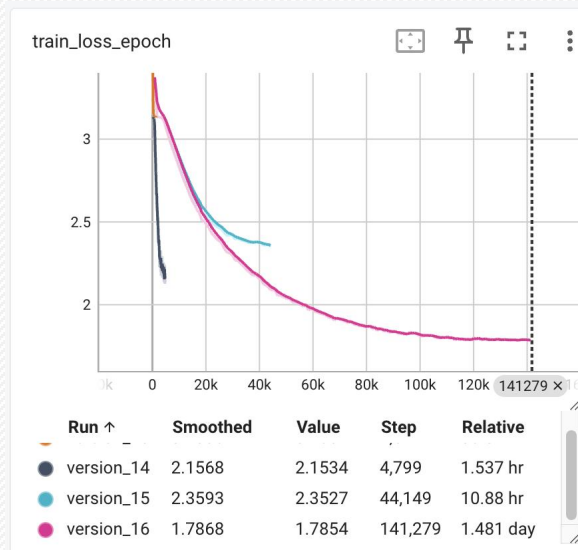
Fine tuning result:

epoch loss: 1.7854;

cross epoch loss : 1.0624;

acc top 1: 0.6111;

acc top 5: 0.9585



04 Finetune (Embedding Model)

Experiment 3 Demo

跨模态的向量相似性检索
Vector Similarity Search Across Audio, Video, Text, Image based on MLLM

by Cuijuan(Jada) Xia2024-10-19 16:11:32

Choose a model

Input a text query

Input an image query

将图像拖放到此处
-或-
点击上传

Clear Submit

Text

Standard Image

Class

Description

Flag



04 Finetune (Chat Model) with RDF data as instruction prompts

Fine tuning experiment 4:

Model size: 2.8B

Hardware: NVIDIA 4090D 24G

Dataset: 277(96+181) GuQin Subtractive

Notation with 277 images; 22 rounds of dialogues for each Notation, as instruction prompts transformed from RDF data.

Fine tuning time: 1000 steps/0.5 hour

Hyperparameter setting: lr 1e-5

Fine tuning results:

train_loss : 1.1058

```
"conversations": [
  {
    "role": "user",
    "content": "这是一个古琴减字谱",
  },
  {
    "role": "assistant",
    "content": "这个谱字的左手所用手指是什么?"
  },
  {
    "role": "user",
    "content": "这个谱字右手手指所弹的弦数是什么?"
  },
  {
    "role": "assistant",
    "content": "这个古琴减字谱谱字的右手手指所弹的弦数是: 四弦。"
  },
  {
    "role": "user",
    "content": "请解释这个古琴减字谱谱字的含义或演奏指法。"
  },
  {
    "role": "assistant",
    "content": "这个古琴减字谱谱字的含义或演奏指法如下: 用左手大拇指按住八徽五分的同时, 用右手大拇指由内向外擘四弦。"
  },
  {
    "role": "user",
    "content": "请解释得详细一点。"
  },
  {
    "role": "assistant",
    "content": "这个古琴减字谱谱字的含义或演奏指法如下: “大八半擘四”是古琴减字谱中的一个指法, 需要左右手同时配合弹奏。“大”代表左手大拇指, “八”代表左手八徽五分, “半”代表左手八徽五分和九徽之间的位置, “擘四”代表右手大拇指由内向外擘四弦。"
  },
  {
    "role": "user",
    "content": "这个谱字的右手所用手指是什么?"
  },
  {
    "role": "assistant",
    "content": "这个古琴减字谱谱字的右手所用手指是: 大拇指。"
  }
]
```

04 Finetune (Embedding Model)

Experiment 4 Demo

古琴减字谱多模态大模型
(GuQin_MiniCPM-V-LoRA)
对话演示

max_length: 2048 (range 0 to 4096)

top_p: 0.80 (range 0.00 to 1.00)

temperature: 0.70 (range 0.00 to 1.00)

Clear chat history

Select mode: Image

Upload image: Drag and drop file here (Limit 200MB per file • JPG, JPEG, PNG)

Browse files

跪七挑一.png (9.2KB)

Deploy

User uploaded image (四七)

User uploaded image (七)

Enter your question



Future work

- Import audio modality into the training dataset
- Expand training dataset scale to 27000+ GuQin Subtractive Notation with 100000+images to enhance the embedding model
- Try the Ontology-driven RAG/GraphRAG experiments to enhance the chat model
- Try the continue pre-training method (full parameters update)
- ◆ Launch a web demo for internal testing by experts (end of October 2024)
- ◆ Launch a web demo for opening testing by the public (end of 2024)
- ◆ Chat Model open source on github (end of 2024)
- ◆ Self-made training dataset open source on github (mid-2025)

Thanks

Xia Cuijuan

xtykc@yeah.net