

CHATGPT

在图书情报界的应用与展望

林 夏

美国 宾夕法尼亚州 费城
德雷塞尔大学 计算信息学院

June 17, 2023



DREXEL UNIVERSITY

College of

Computing & Informatics

回顾：图书馆该如何欢迎数字人到馆服务

- ADLS 2022
- Digital Humans – 数字人与图书馆的将来
 - Conversational Interfaces
 - Personalized Service
 - Knowledge Service
 - Digital Twin Service
 - Digital Humans in Libraries



数字人与图书馆

- 数字人将如何改变图书馆的服务
 - 更人性化的服务
 - 更个性化的服务
 - 更深度化的知识服务
 - 数字人为数字人提供图书馆服务
- CHATGPT会是这样一个数字人吗？
 - 看上去不像！
 - 说起来有点像？
 - 有知识，有智商，直得培养！



ChatGPT 与 数字人

- “有趣的灵魂”与“好看的皮囊”
 - 创造数字人图书馆员来取代图书馆员？
 - Everything you learn in LIS schools, chatGPT knows
 - Any question you ask, chatGPT has answers.
 - 创造智能助手来帮助图书馆员的工作？
 - to complete 80% of intellectual work quickly
 - to help experts complete the rest of 20% (24 hours a day)

将来？

现在



DREXEL UNIVERSITY

College of

Computing & Informatics

ChatGPT的智慧在哪里？

- 知识庸博
- 夸夸其谈
- 浮想联翩
- 知无不言
- 胡编乱造





中国图书馆的未来充满了潜力和机遇。随着数字化时代的到来，图书馆正在逐步转变成为知识服务中心，不仅提供传统的纸质书籍和阅览空间，还提供数字化资源和创新的学习环境。



以下是一些可能塑造中国图书馆未来的趋势和发展方向：

1. 数字化转型：随着数字化资源的广泛发展，图书馆将继续数字化转型，提供电子书籍、期刊、数据库等数字资源，以满足用户在线阅读和学习的需求。数字化还将促进图书馆间的合作与资源共享。
2. 技术创新：图书馆将积极应用先进技术，如人工智能、大数据分析和虚拟现实等，来改进用户体验、提供个性化服务和推动知识获取的方式。这些技术可以用于智能图书推荐、学习辅助、社交互动等方面。
3. 学习与创新空间：图书馆将进一步发展成为学习和创新的中心，提供多功能的学习空间、创客空间、演讲厅和会议室等，以满足不同用户的学习和交流需求。这些空间可以支持小型研讨会、创业活动、技能培训等。
4. 社区参与：图书馆将继续与社区紧密合作，成为社区的文化和教育中心。通过组织社区活动、阅读俱乐部、讲座和展览等，吸引更多的人参与到图书馆的活动中来，促进文化交流和社会发展。
5. 跨界合作：图书馆将与其他机构和组织进行更多的跨界合作，如学校、博物馆、科研机构和企业等。通过合作，图书馆可

↻ Regenerate response



ChatGPT弱智在哪里？

- 知识不问出处
- 道德底线不清



ChatGPT与扫地僧

- 有机会接触所有的知识
- 无限制，无约束的学习
- 无人所知的超强“武功”
- 无意识使用“武功”
 - 用的上时武力超强！
 - 用不上时“继续扫地”
- 武术流派的最后防线
 - 人类最后的机会？



雇佣chatGPT 到图书馆

- 应该考虑的问题
 - 不是它会做什么，而是工作最需要的是什么！
- 在图情工作中，哪些是最耗费人力与智能的？
 - Generating metadata (元数据自动生成)
 - Building controlled vocabularies (构建专业主题词汇表)
 - Taxonomy development (专业分类词汇表的开发)
 - Creating semantic links/automatic indexing
 - Creating Knowledge Space



Taxonomy Development

- 分类词汇表的开发
 - Define a scope – 词汇表的目的，用途，使用对象，方式等
 - Create – 通过有系统有组织的分类来收集词汇，定义词汇
 - Implement – 根据专家的意见决定更详尽的分类与分类的层次，词汇等
 - Test – 在测试中修改完善分类词汇表
 - Maintain – 在使用中更新完善



案例：深度学习的分类

4. Deep Learning

- 4.1 Convolutional Neural Networks (CNNs)
 - 4.1.1 Image Classification
 - 4.1.2 Object Detection
 - 4.1.3 Image Segmentation
 - 4.1.4 Style Transfer
 - 4.1.5 Generative Adversarial Networks (GANs)
- 4.2 Recurrent Neural Networks (RNNs)
 - 4.2.1 Natural Language Processing (NLP)
 - 4.2.2 Speech Recognition
 - 4.2.3 Time Series Analysis
 - 4.2.4 Sequence-to-Sequence Learning
 - 4.2.5 Long Short-Term Memory (LSTM)
- 4.3 Autoencoders
 - 4.3.1 Anomaly Detection
 - 4.3.2 Dimensionality Reduction
 - 4.3.3 Data Compression
 - 4.3.4 Image Denoising
 - 4.3.5 Generative Models
- 4.4 Reinforcement Learning
 - 4.4.1 Game Playing
 - 4.4.2 Robotics
 - 4.4.3 Control Systems
 - 4.4.4 Decision Making
 - 4.4.5 Multi-Agent Systems



用CHATGPT 助分类词汇表的开发

- 步骤：
 - Describe the scope and first level of Taxonomy to chatGPT
 - Expand individual categories to deepen the taxonomy hierarchy.
 - Request chatGPT to define the terms or categories.
 - Remove or modify terms generated by chatGPT
 - Experts review and refine the taxonomy



Learn to speak chatGPT

- Example: “Expand categories of a taxonomy for data analytics consultants working with small business:
 - 1. Data Collection
 - 2. Data Analysis
 - 3. Business Strategy
 - 4. Performance Measurement
 - 5. Project Implementation
 - 6. Training and Education”

3.1. Business Strategy

- Creating Data-Driven Business Strategies
- Identifying Key Performance Indicators (KPIs)
- Business Process Improvement
- Risk Analysis and Management
- Market Research and Competitive Analysis



Refine the taxonomy

< 2/2 >

XL

Create subcategories for the above category "Business Process Improvement" using nouns only

对“Business Process Improvement”继续分类



Sure, here are some subcategories for "Business Process Improvement" using only nouns:



1. Workflow Optimization
2. Process Mapping
3. Bottleneck Identification
4. Root Cause Analysis
5. Process Standardization
6. Process Automation
7. Quality Control
8. Lean Six Sigma
9. Process Metrics
10. Continuous Improvement



DREXEL UNIVERSITY

College of

Computing & Informatics

Create Interview Questions

XL

Generate interview questions for the consultant to talk with her business client on the category of "Business Process Improvement" and "Root Cause Analysis"

生成与“Business Process Improvement” and “root Cause Analysis”有关的咨询问题



Sure, here are some in business client regard Cause Analysis":

Business Process Imp

1. What are some of t
2. How do you current processes?
3. Have you ever conc improvement?
4. What are some con processes?
5. Have you ever cons streamline your bus
6. How do you ensure your organization?

Root Cause Analysis

1. How do you currently investigate and address issues or problems within your organization?
2. What are some of the most common types of issues or problems that you encounter?
3. How do you currently identify the root causes of issues or problems within your organization?
4. Have you ever conducted a formal root cause analysis exercise?
5. What tools or techniques do you currently use to identify and address the root causes of issues or problems?
6. How do you prioritize which issues or problems to address first?
7. How do you ensure that solutions to issues or problems are effective and sustainable over time?
8. How do you involve employees and other stakeholders in the root cause analysis and problem-solving process?
9. Have you ever experienced any challenges with implementing solutions to address root causes?
10. How do you measure the success of solutions implemented to address root causes?



DREXEL UNIVERSITY

College of

Computing & Informatics

Create a ChatGPT Plug-in for Taxonomy Creation

Using chatGPT's APIs:

```
while(True):
    question=input('user question: ')
    user_input = text + question
    conversation.append({"role": "user", "content": user_input})
    conv_history_tokens = num_tokens_from_messages(conversation)

    while (conv_history_tokens+max_response_tokens >= token_limit):
        del conversation[1]
        conv_history_tokens = num_tokens_from_messages(conversation)

    response = openai.ChatCompletion.create(
        model="gpt-3.5-turbo-0301",
        messages = conversation,
        temperature=0,
        max_tokens=max_response_tokens
    )

    conversation.append({"role": "assistant", "content": response['choices'][0]['message']['content']})
    print("\n" + question + "\n" + response['choices'][0]['message']['content'] + "\n")

    if question=='print final taxonomy categories':
        with open(fwrite,'w') as f:
            f.write(response['choices'][0]['message']['content'])
        print('stop')
        break
```



Create a ChatGPT Plug-in for Taxonomy Management

- Major Functions:
 - Read in the hierarchy of a taxonomy from a file.
 - Expand a category or every categories in a branch.
 - Remove a category or a sub-category
 - Replace a category with human editing
 - Save current taxonomy to a file.
- Others:
 - Create definitions for the category terms.
 - Generate interview questions based on categories.
 - Create relationships of the terms?



Next Project : Ontology 对比与测试

• Developing/Expanding Existing Ontology

Chemical Process (A)

- ☰ 🔍 - Chemical Property
 - a. Acidity
 - b. Alkalic
 - c. Alkaline
 - d. Alkalinity
 - e. Basic
 - f. Basicity
 - CarbonateAlkalinity
 - TotalAlkalinity
 - g. Composition
 - Abundance
 - IonComposition
 - h. Purity
- ⊙ Chemical Quantity
- ⊙ Component
 - Consistence Property
- ⊙ Density
 - Dissolved Solid
- ⊙ Dry Weight
- ⊙ Mass Concentration
- ⊙ Property
- ⊙ Reaction Property

Chemical Process (B)

- Chemical Property
 - a. Acidic property
 - b. Basic property
 - c. Redox property
 - d. Solubility property
 - e. Chemical reactivity property
 - f. Heat of combustion property
 - g. Flammability property
 - h. Toxicity property
 - i. Radioactivity property
- ⊙ Chemical Quantity
- ⊙ Component
- ⊙ Consistence Property
- ⊙ Density
- ⊙ Dissolved Solid

Which one is the ontology published in BioPortal?
Which one is generated by chatGPT?



Developing/Expanding Existing Ontology

- Density
 - ◉ Load
 - ◉ OxygenDemand
 - TotalDissolvedSolids
 - TotalSuspendedSolids
- ◉ Dissolved Solid
- ◉ Dry Weight
- Mass Concentration
 - Salinity
 - Sodicity
 - Water Concentration
 - Moisture

- Density
 - a. Bulk density
 - b. Particle density
 - c. Specific gravity
- ◉ Dissolved Solid
- ◉ Dry Weight
- Mass Concentration
 - a. Parts per million (ppm)
 - b. Parts per billion (ppb)
 - c. Grams per liter (g/L)
 - d. Milligrams per liter (mg/L)

Which ontological relationship make sense to you for the “Density” class and “Mass Concentration” class?



Developing/Expanding Existing Ontology

- Observations:
 - Human developed ontology
 - tends to be bias to individual experts’ “ontological views”
 - Incomplete, imbalanced classes
 - Specific to the tasks that ontology was developed for
 - ChatGPT assisted ontology development
 - More comprehensive (due to its large language models)
 - More general and lack of specifics
 - Classification rather than ontological approach



chatGPT in Libraries

- 深度化的咨询服务
 - Library users expect they can talk to “librarians” any time on any topic.
- 个性化的信息服务
 - Access, digest, recommend, **synthesize** INFORMATION based on what individual knows
 - Create a personal librarian for every library user
 - 为每一个用户培养一个“个人图书馆员”
- 高效率的知识服务
 - Create taxonomy, ontology, and personalized knowledge space for users



chatGPT in Libraries

- Prompt Engineering
 - developing and optimizing prompts to efficiently use language models (LMs) for a wide variety of applications and research topics
- Prompt Engineering in Libraries
 - 学会与chatGPT对话
 - 学会与AI对话
 - 帮助图书馆用户与chatGPT对话
 - 帮助图书馆用户与AI对话



结论

- CHATGPT将是我们的高级智能助手
 - 分担智能工作
 - 开拓我们的知识范畴
 - 缩短项目开发时间
- 要学会如何引导ChatGPT分担我们的工作
 - ChatGPT是一种语言
 - 学会沟通，效率百倍
 - 应该通过智能助手ChatGPT提供个性化与高效率的知识服务。

