


# Semantic Web Development in Traditional Chinese Medicine

Huajun Chen  
Zhejiang University



# Semantic Web Development for Traditional Chinese Medicine

What's TCM?

TCM Semantic Web

TCM Ontology Engineering

TCM Semantic Search Engine (DartGrid System)

Semantic Graph Mining for biomedical network analysis

# What's TCM

---

# What's TCM?

- Traditional Chinese Medicine (TCM) is an ancient medical system that accounts for around 40% of all health cares delivered in China.
- Preventive Medicine
  - Take medicine as like a daily nutrition supplement or part of food to maintain the balance of the whole body system.
- Personalized Medicine
  - Treatment can be completely different for people with respect to their gender, age, health condition although they have very similar symptoms.
- Empirical Medicine
  - The effect of many TCM drugs are based on more one thousand years of practices, whereas they do not know the specific underlying mechanism.

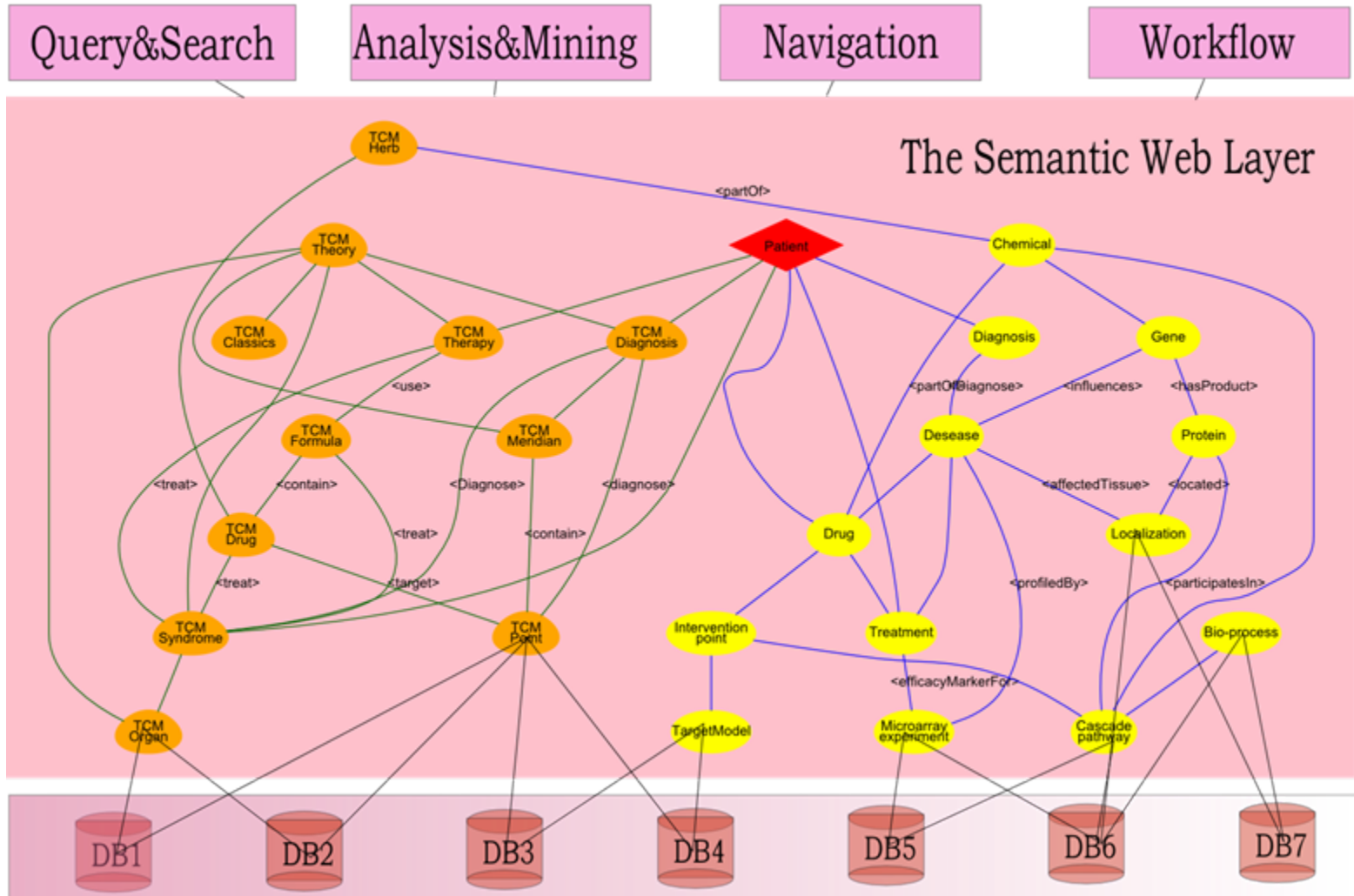
# TCM Knowledge

- TCM theories derive from many knowledge sources including the theories of Yin-Yang, Chinese five elements, the human body channel system, Zang Fu organ theory, holistic connections, mind-body intervention, and many others.
- TCM practice includes diagnosis and treatments theories such as herbal medicine and , massage and cupping, acupuncture and meridians.

# TCM Semantic Web Project

A project in collaboration with China Academy of Traditional Chinese Medicine.

# The ultimate vision of the TCM Semantic Web



# The Subprojects

- TCM Ontology Engineering (2001-current).
- The DartGrid Data Integration System (first started in 2002)
- Integrating legacy relational database into Semantic Web
  - DartMapper: Visualized relational-2-RDF Mapper (2003-2005)
  - DartQuery: SPARQL2SQL Query Rewriter and a Form-based SPARQL query builder (2003-2006)
  - DartSearch: Semantic Search. (2005-current)
- Semantic Data Analysis and Data Mining for Semantic Web
- DartSpora: semantic data analysis engine (2007-current)
- Semantic Graph Mining for biomedical network analysis. (2007-current)

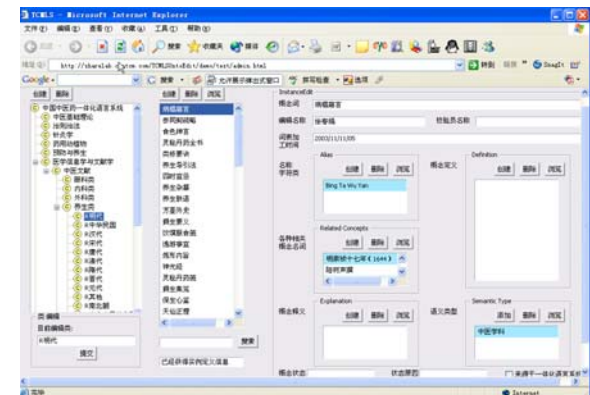


# TCM Ontology Engineering

---

# TCM Ontology Engineering

- A effort participated by more than 100 persons from over 30 TCM research institutes located in different parts of China
- Scale
  - More than 20,000 classes and 100,000 instances defined in the current ontology
- Service  
Web APIs for ontology-based applications.



The current TCM ontology contains **15 major categories** for each sub-domain.

Categories	Concepts	Individuals
Pathogeny and pathogenesis and diagnosis	65	4939
Geography	948	86
Pharmacology of TCM medical formulae	158	11677
Disease	308	6516
Humanities	1680	10448
Health care management	303	1300
Medicinal plant and animal	719	14804
Informatics and philology	216	21089
Prevention and life-cultivation	163	69
Acupuncture	716	1647
Therapeutic principles and treatments	108	1997
Traditional Chinese Drugs	1572	3624
The basic theory of TCM	3313	1725
The doctrines of TCM and relevant sciences	1352	219
Natural science and physics	217	240

# Ontology visualization and query engine

设定展示区域 | 浏览在线帮助 首页 登录

欢迎使用 - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 搜索 收藏夹

地址 http://localhost:8080/DartWiki/Wiki.html#%E5%A4%A7%E9%BB%84 转到 链接

设定展示区域 浏览在线帮助 首页 登录

## 中医百科

大黄     模糊

基本信息 中药

被...使用 [论利湿清热 \(中医学语言系统\)](#)

被...使用 [论舌胀大不能出口而神情清爽 \(中医学语言系统\)](#)

被...使用 [论实热吐血 \(中医学语言系统\)](#)

上位词 [攻下药 \(中医学语言系统\)](#)

有---的概念部分 [《与疫新论》 \(中医学语言系统\)](#)

被---治疗 [论邪滞津枯证 \(中医学语言系统\)](#)

有---的概念部分 [《十三制大黄经验良方》 \(中医学语言系统\)](#)

被包含 [论寒燥 \(中医学语言系统\)](#)

被包含 [创三消饮 \(中医学语言系统\)](#)

is a [攻下药 \(中医学语言系统\)](#)

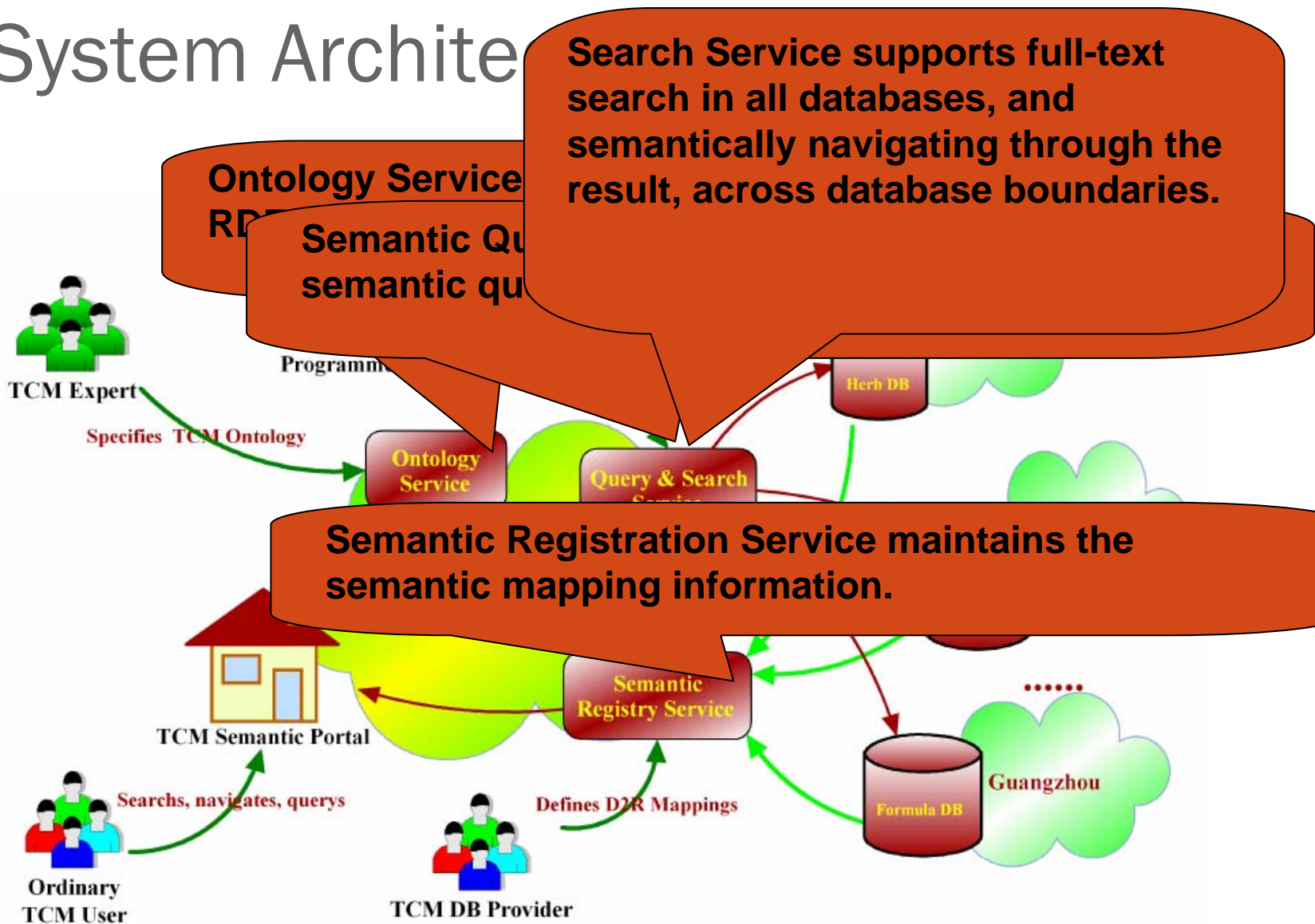
有---的概念部分 [《辨疫琐言》 \(中医学语言系统\)](#)

本地 Intranet

# TCM Semantic Search Engine

A semantic search engine build upon a lot of relational databases.

# System Architecture



# Visualized Mapper

1. Display relational tables

2. Display ontologies.

3. User drags tables and classes into this panel, and establishes their mappings. One table is likely to be mapped to more than one classes.

4. (unlabeled)

5. Meta-information about the selected table.

6. Outline of the mapping definitions. User could query mappings defined before.

**DbRes Tree View**

- {http://dart.zju.edu.cn/tcm/}
- {http://dart.zju.edu.cn/tcm/}
- JMZ\_TMZ\_TZT
- TCM\_BASIC.ZT\_VIEW\_ANMOLF
  - AMCS
  - AMFF
  - AMSF
  - AMSJ
  - JL
  - XW\_ID
  - XWLX
  - XWMC

**OntoSchema Tree View**

- 按摩疗法
  - 按摩次数
  - 按摩方法
  - 按摩时间
  - 按摩手法
  - 经络
  - 临床治疗步骤序号
  - 穴位名称
  - 治疗方法ID
  - 治疗类型
  - 治疗时间
  - 治则
- 病毒性肝炎

**Workspace**

**ZT\_VIEW\_ZHENJIULF**

- Ontology: http://dart.zju.edu.cn/dartcore/dart#针灸疗法
- Property: http://dart.zju.edu.cn/dartcore/dart#针刺次数
  - Column: ZCCS
- Property: http://dart.zju.edu.cn/dartcore/dart#针刺手法

**ZT\_VIEW\_ANMOLF**

- Ontology: http://dart.zju.edu.cn/dartcore/dart#按摩疗法
- Property: http://dart.zju.edu.cn/dartcore/dart#按摩方法
  - Column: AMSF

**Properties**

Property	Value
Db Localpart	TcmBasic
Db Namespace	http://dart.zju.edu.cn/tcm/database
Table Name	ZT_VIEW_ZHENJIULF
Table Schema	

**Outline**

- Table: ZT\_VIEW\_ZHENJIULF
  - Ontology: http://dart.zju.edu.cn/dartcore/dart#针灸疗法
  - Property: http://dart.zju.edu.cn/dartcore/dart#针刺次数
  - Property: http://dart.zju.edu.cn/dartcore/dart#针刺手法
- Table: ZT\_VIEW\_ANMOLF
  - Ontology: http://dart.zju.edu.cn/dartcore/dart#按摩疗法
  - Property: http://dart.zju.edu.cn/dartcore/dart#按摩方法
- Table: JMZ\_LCYJ

# Semantic Search Portal Version 1

The image displays a screenshot of a Semantic Search Portal in a Microsoft Internet Explorer browser. The interface is annotated with several callout boxes and a central diagram.

**Callout Box 1 (Left):** Labeled "Ontological classes", it points to a list of categories such as "方剂" (Prescriptions), "中药" (Herbs), and "TCM\_ADMIN ZYFU".

**Callout Box 2 (Bottom Left):** Explains that when full text search returns too many results, clicking the classes leads to a dynamic form-based query interface for more accurate results.

**Callout Box 3 (Top Center):** Labeled "动态网页的基本信息" (Basic information of dynamic web pages), it points to a table listing various items with columns for 序号 (Serial Number), 类别名称 (Category Name), 选择功能 (Selection Function), 选择操作 (Selection Operation), 选择内容 (Selection Content), and 标准网页提示 (Standard Web Page Prompt).

**Callout Box 4 (Bottom Center):** States that based on the semantic relations defined at the ontological level, users can keep searching and navigating over integrated databases without awareness of database boundaries.

**Central Diagram:** A hierarchical structure showing "Synonyms and Paronyms" leading to "Semantic association", which then branches into various search results like "经方", "方剂", "中药", "医案", etc.

序号	类别名称	选择功能	选择操作	选择内容	标准网页提示
1	经方	<input type="checkbox"/>	包含		提示 译地址...
2	方剂应用	<input type="checkbox"/>	包含	耳部常见病	提示 耳部常见病
3	经方	<input type="checkbox"/>	包含	脚气	提示 脚气
4	方剂	<input type="checkbox"/>	包含		提示 译地址...
5	方剂	<input type="checkbox"/>	包含		提示 译地址...
6	方剂	<input type="checkbox"/>	包含		提示
7	方剂	<input type="checkbox"/>	包含		提示
8	方剂	<input type="checkbox"/>	包含		提示
9	方剂	<input type="checkbox"/>	包含		提示



# Semantic Search Portal Version 1

http://localhost:8080/mv/Search/index.html - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

地址(O) http://localhost:8080/mv/Search/index.html#

数据库 网页 图片 论文 百科 挖掘

大黄 中医药搜索

共搜索到25454条记录

1... << < 1 2 3 4 5 6 7 8 9 10 > >> ...2546 跳至

**题目: 大黄**

作者: 黄煌  
摘要:  
刊名: 中国社区医师  
年份: 2002

**题目: 大黄与藏边大黄显微定量鉴别**

作者: 李军林;李家实;王爱芹  
摘要: 利用**大黄**及其易混品藏边**大黄**粉末中草酸钙簇晶数目差异, 鉴别**大黄**与藏边大 黄, 方法简便, 准确性高。  
刊名: 中药材  
年份: 1996

**题目: HPLC测定枝穗大黄中大黄素和大黄酚的含量**

作者: 常军民, 高宏, 张焯, 赵军, 堵年生  
摘要: 目的:测定枝穗**大黄**中**大黄素**和**大黄酚**的含量。方法:采用HPLC测定。结果:枝穗**大黄**中**大黄素**和**大黄酚**的平均含量分别为0.04%和0.23%;加样回收率**大黄素**为99.2%,**大黄酚**为100.7%。结论:所用方法快速、灵敏、准确, 样品处理简单。  
刊名: 华西药理学杂志  
年份: 2003

**题目: 大黄的外用经验**

作者: 梁后权  
摘要: 介绍了国内期刊报道的**大黄**外用及**大黄**煎剂保留灌肠治疗17种疾病的经验。  
刊名: 贵阳中医学院学报  
年份: 2002

相关搜索

针灸

其他

中医理论

化学物质

草...

完毕

# Semantic Data Analysis for TCM

What kinds of new connections can be discovered or mined from this huge web of data?

# Graph vs Semantic Graph

Conventional Graph

Semantic Graph Model

Node

Semantic Graph as a Knowledge Base

Semantic Graph as a Complex Network

different

Edge

All

labeled, different

Basic Element

No

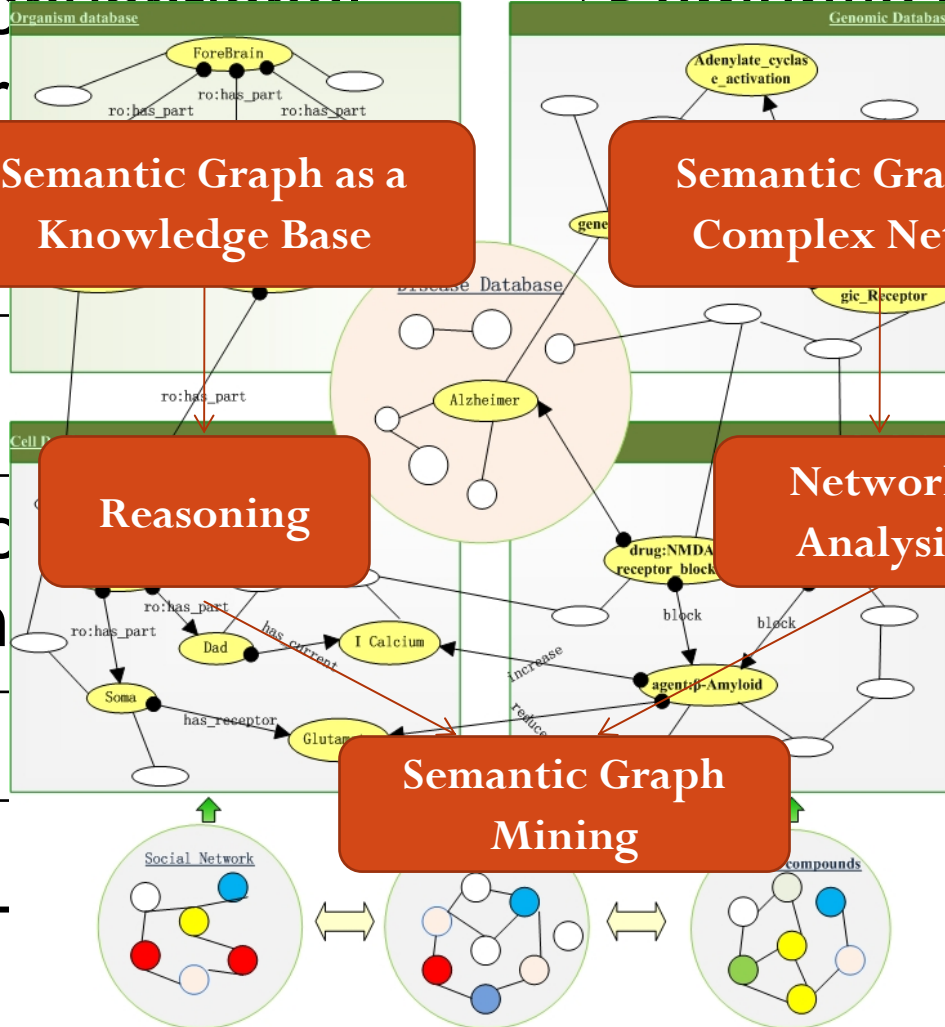
Reasoning

Network Analysis

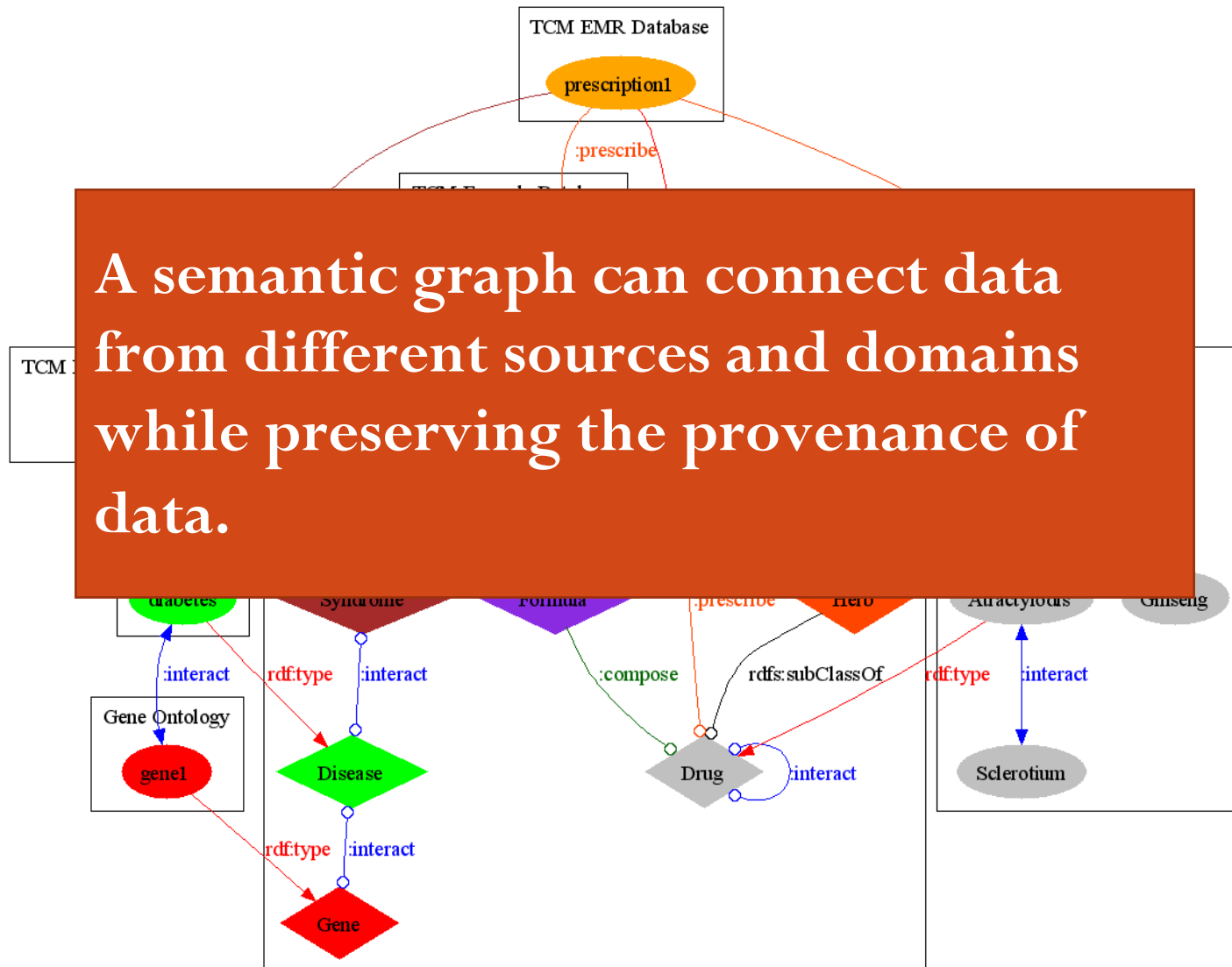
different stands

Element

Semantic Graph Mining

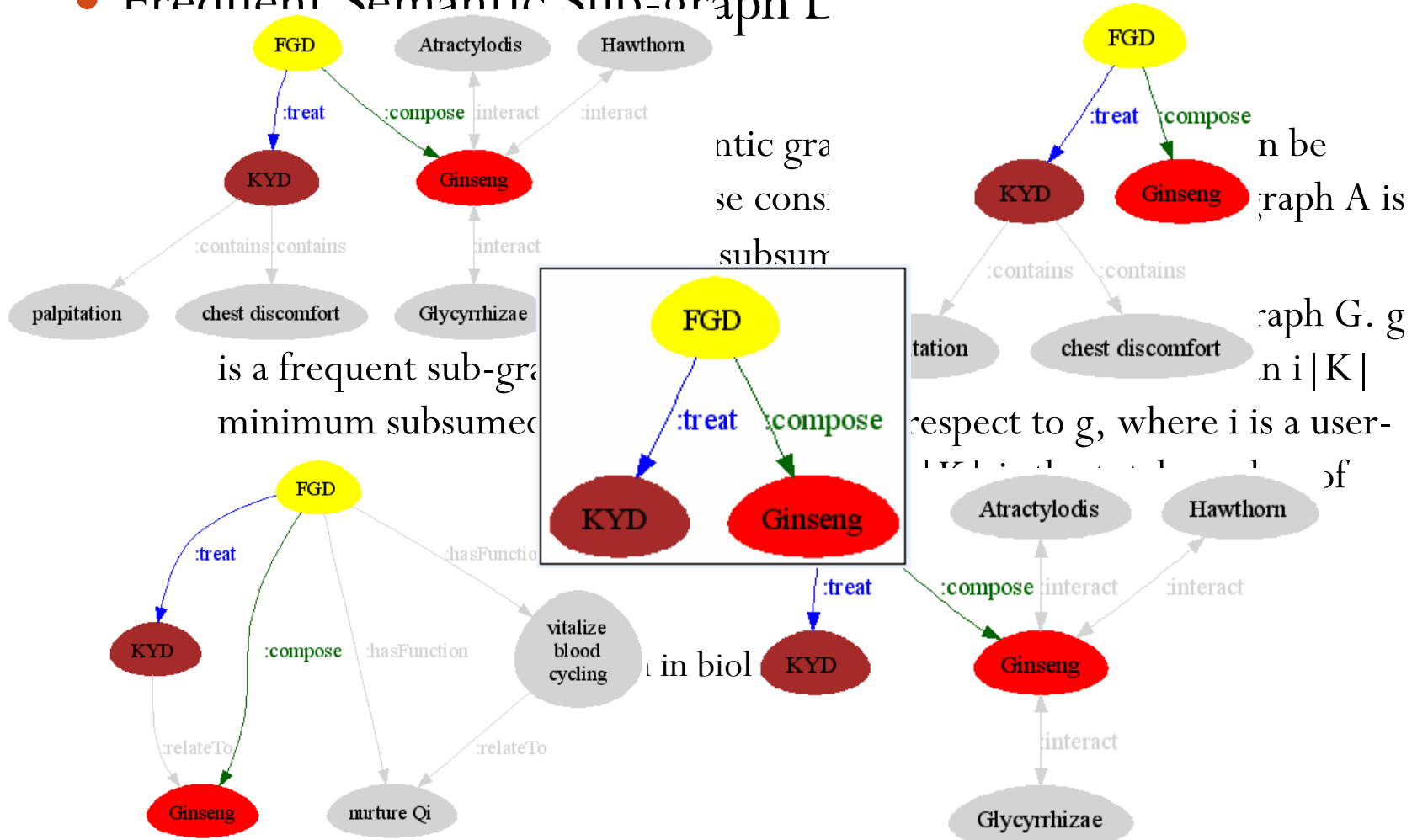


# An example.



# An example

- Frequent Semantic Sub-graph  $\Gamma$ :



# Semantic data analysis

- Semantic graph contains richer information than normal graph.
- It is based upon the integration capability of semantic web.
- Much more meaningful mining results:
  - Discover the facts directly.
  - Find more meaningful associations among entities.
  - Calculate the network parameters in a more accurate way.
- Ontological reasoning can be leveraged to further facilitate the mining process.
- We need good tools to help do so.

# DartSpora: a interactive mining engine for TCM

The screenshot displays the Spora Data Mining Platform interface. At the top, the browser address bar shows the URL `http://localhost:8888/cn.edu.zju.ccnt.dartkd.spora.Spora/Spora.html`. The main title is "Spora 中医药数据挖掘平台". Below the title, there are navigation tabs: "About Spora", "Operator", "Operator Parameter", "Experiment Result", and "Login Control".

The interface is divided into several sections:

- OperatorTree Explorer:** A tree view on the left showing the hierarchy of operators. The "TCM" folder is expanded, showing operators like "SplitComponent", "PreProcessUnderRules", "SimpleSplitor", "PreProcessByRegularExp", "MaxPatternOperator", "NotNullColumnFilter", "MappingStrTold", and "MappingIdToStr".
- Experiment Explorer:** A section on the left showing a tree view of experiment components, including "PreprocessByRegExp", "Root", "ExcelExampleSource", "PreProcessByRegularExp", "NotNullColumnFilter", "MappingStrTold", "MappingStrToI", "MaxPatternOperator", and "MappingIdToStr".
- Results:** A table in the center-right showing the results of experiments. The table has columns for "Name", "FWEC", "SubMethodations", "Weight", "Unit", and "ChChi".
- Data Visualizations:** Several charts are displayed, including a pie chart, a bar chart, and a network graph.

At the bottom of the interface, it says "DartKD Team @ CCNT Lab Zhejiang U".

# Summary

- A Web of Data means a lot to us.
- It can enable fancy ways of searching and browsing the daunting online information space.
- It can also finally unleash the potential underlying disparate data sources to greatly facilitate and advance the data mining and knowledge discovery technology.
- But we need powerful tools to help us to achieve the goal.



# Summary:

## Key Benefits of Semantic Web for TCM

- Fusion of data across many scientific discipline
- Easier recombination of data
- Querying of data at different levels of granularity
- Capture provenance of data through annotation
- Data can be assessed for inconsistencies
- Integrative knowledge discovery from large-scale semantic graph formed by integrating cross-institutional, cross-disciplinaries data sources.

Thanks for your time!