**第一章\_德塔自然语言图灵系统**

**测试速度：**单机联想Y7000笔记本win10 实测峰值每秒 中文分词1630~1650万+中文字， 词库65000+，函数准确率100%，缺失语法函数 0.3%-， 算法准确率 99.7%+， 100%完整开放源码，在api与书籍中。

**测试效果：输入：**如果从容易开始于是从容不迫天下等于是非常识时务必为俊杰沿海南方向逃跑他说的确实在理结婚的和尚未结婚的提高产品质量中外科学名著内科学是临床医学的基础内科学作为临床医学的基础学科重点论述人体各个系统各种疾病的病因发病机制临床表现诊断治疗与预防

**输出结果：**如果+从+容易+开始+于是+从容不迫+天下+等于+是非+常识+时务+必+为+俊杰+沿海+南+方向+逃跑+他+说+的+确实+在理+结婚+的+和+尚未+结婚+的+提高+产品质量+中外+科学+名著+内科学+是+临床+医学+的+基础+内科学+作为+临床+医学+的+基础+学科+重点+论述+人体+各个+系+统+各种+疾病+的+病因+发病+机制+临床+表现+诊断+治疗+与+预防+++++



定义：德塔分词是一种-- 基于神经网络索引字典切割-- 进行前序遍历词性组合匹配-- 按文学语法定义搭配 的切词引擎。

**德塔分词的催化切词优化方式主要包含：**

1 索引字典进行细化拆分加速。

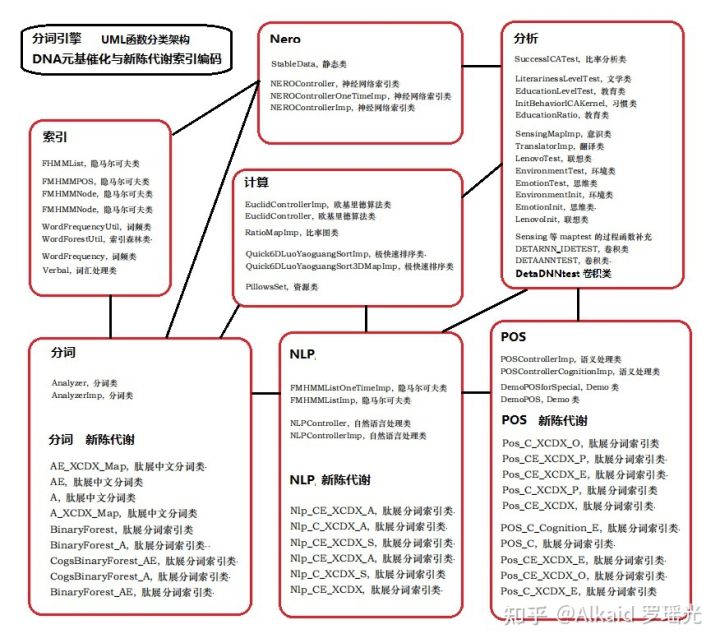
2 函数进行使用频率统计排列加速优化。

3 动态类卷积遍历内核的关键字优化。

4 函数文件和 函数文件名 进行新陈代谢，二次新陈代谢优化索引编码加速。

5 文学切词语法函数的细化优化加速。

定义者 罗瑶光



**分词，**

1 德塔的分词是一种前序《排队论》逐字遍历文字索引，通过索引中的词汇匹配 按长度进行提取，然后将提取的词汇串 进行词性切分的过程。refer page 12 ~

2 德塔的分词文字索引采用关联分类生成小文件map集（词性map，词长map，词类map）， 进行整体加速，作为一个催化细化过程。refer page 44，54, 92，

3 德塔的词汇匹配目前有多个国家语言字符集，可统一，可拆分，目前最大划分处理长度为4，划分切词采用动态 类似CNN 卷积（遍历pos函数语句的内核计算，非卷积的积分叠加计算） StringBuilder核做POS识别。refer page 45，119,120，

4 德塔的词性切分按照4字词 3字词 2字词 单字 进行逐级按词汇的 POS搭配语法模式进行归纳，按文本的POS出现频率进行流水阀门方式优化。refer page 97,116，



（德塔分词逻辑， 已经纠正红色字 ‘卷积’改为‘内核’，因为第四修订版本已经在申请中，ppt所有书中的原图纠正内容统一更新在第5版，罗瑶光）

**排序，**

1 德塔分词排序思想原型采用 Sir Charles Antony Richard Hoare 的 快速排序思想。

refer page 版权原因无文字收录 已经refer [快速排序算法\_百度百科](https://link.zhihu.com/?target=https://baike.baidu.com/item/%E5%BF%AB%E9%80%9F%E6%8E%92%E5%BA%8F%E7%AE%97%E6%B3%95/369842?fr=aladdin" \t "https://zhuanlan.zhihu.com/p/_blank)

2 德塔分词排序源码原型采用 Introduction to Algorithms 的 快速排序4代源码。

refer page 版权原因无源码收录 已经refer [https://github.com/yaoguangluo/Data\_Processor/blob/master/DP/sortProcessor/Quick\_4D\_Sort.java](https://link.zhihu.com/?target=https://github.com/yaoguangluo/Data_Processor/blob/master/DP/sortProcessor/Quick_4D_Sort.java" \t "https://zhuanlan.zhihu.com/p/_blank)

3 基于1 和 2原型，德塔分词排序 采用 Theory on YAOGUANG's Array Split Peak Defect 的微分催化算子优化思想 2013年开始优化。refer page 247,248,250,529,620，

4 优化过程为 小高峰左右比对法， 波动算子过滤思想，离散条件归纳微分思想(如狄摩根计算，流水阀门计算等)，目前为TopSort5D。refer page 658，下册134

5 德塔分词的函数优化方式和算法优化方式，包括分词引擎，读心术，NLP分析等核心组件均采用 微分催化系统。 refer page 661，

**神经网络索引，**

1 德塔分词的词汇字典用map进行索引，因为jdk8+的map对象的key支持2分搜索，搜索速度到了峰值。refer page,129,131

2 德塔分词的索引不断的将大map进行细化分类，如词长map，词类map，词性map，让搜索再次加速。refer page 55，

3 德塔分词的索引map支持 2次组合计算，支持分布式服务器进行索引cache。关于2次组合计算作者不建议单机使用。refer page 92，

4 德塔分词map的key用string的 char对应ASCII int进行标识来执行find key，方便二分搜索存储和 StringBuilder高速计算，实现底层核统一。refer page 92

**分词在线性文本搜索中应用，**

1 德塔分词的搜索建立在map类的权重计算方法上，不同的权重叠加产生的打分进行排序输出。refer page 下册64

2 权重的计算方法按词性的主谓宾如代 名动形 ，和 POS如 动名形谓介分类。refer page 下册66

3 权重与词长，词频进行耦合bit叠加计算(bit位计算比乘法要快一个数量级)，生成最终输出结果。 refer page 下册68

4 权重与词长的 比值可以精度调节，确定搜索的精确性和记录个人搜索偏好。refer page 下册68

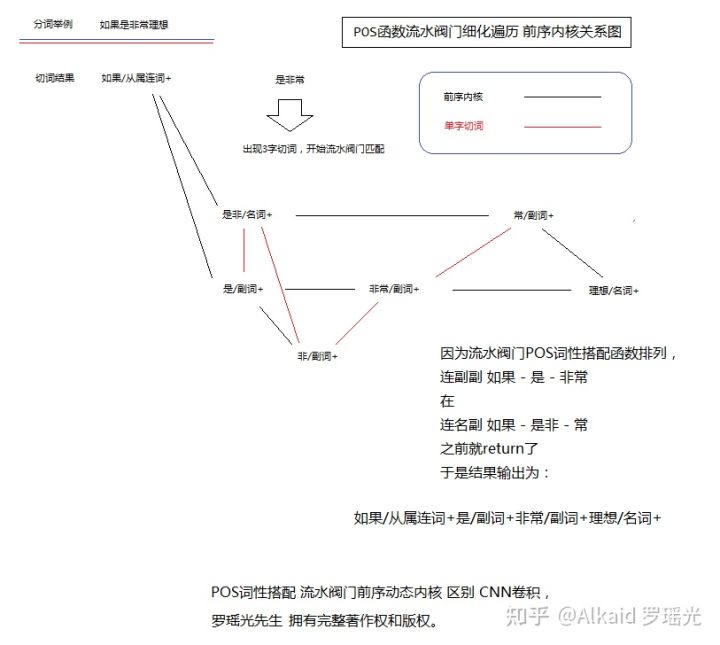
**动态 POS函数流水阀门细化遍历 内核匹配，**

1 动态的核分为前序核和后序核两种。根据词汇分析的位置进行实时变动更新。refer page 97

2 前序核主要缓存存储词汇的位置和词性，用于POS词性搭配的 POS函数流水阀门细化遍历 计算。refer page 97

3 后序核主要缓存词汇的切词链 后面准备 跟进的词语。用于POS语法的修正计算，如连词匹配。refer page 97

4 内核采用StringBuilder做核载体进行计算加速。refer page 97

2019年3月18日之前作者Github的 该算法函数编码框架已经出现

[https://github.com/yaoguangluo/Deta\_Parser/commit/25b90c9847d15df85c5c991448f2c271e0ad8106](https://link.zhihu.com/?target=https://github.com/yaoguangluo/Deta_Parser/commit/25b90c9847d15df85c5c991448f2c271e0ad8106" \t "https://zhuanlan.zhihu.com/p/_blank)

注意：链接的**CNN 关键词**的 历史记录 属于作者用词错误，作者当年基础学术累积不够，关于卷积的知识仅仅学了**计算机视觉**的理论课，以为带内核计算的都叫CNN卷积，

另外作者发现自己还有一个错误， 就是以为序列链表方式计算就叫隐马科夫链计算。所以**CNN+隐马可夫**这两个技术词汇，伴随作者10年之久。今天进行ppt严谨定义，翻阅大量定义文献资料，才发现这些错误。予以纠正。作者的ANN和RNN 出现的文本分析内核计算才是真正的CNN卷积计算。

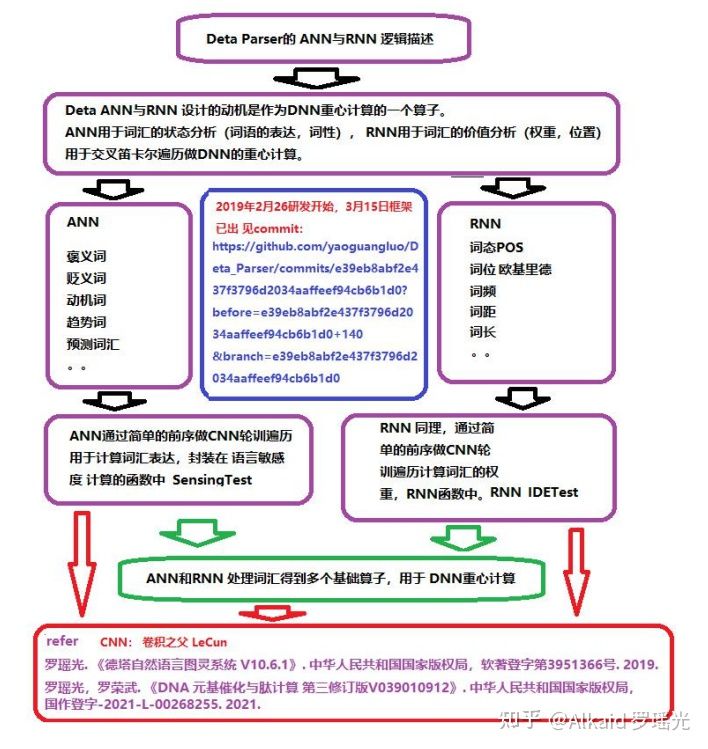
**POS,**

1 德塔分词的核心类，包含了词性的搭配切分所有函数。refer page 97,116



**NLP,**

1 德塔分词的核心类，包含了词性的词长切分所有函数。refer page 119，120



**ANN,**

**德塔词性的卷积计算ANN，主要包含意识比率算子,环境比率算子,动机比率算子,情绪比率算子。这个四个算子 的组合计算产生了一些高级决策,如 情感比重,动机比重,词权比重,持续度,趋势比重,预测比重,猜想比重,意识综合。这些决策在文本分析的领域可以拥有实际评估和决策的价值。同时意识综合 summing 也是德塔DNN计算的一个输入参数组件，用于文本中心思想词汇标识计算。**

1词性卷积计算refer page 182

2用于确定文本的中心

2.1 算子组成

2.1.1 S SENSING 意识比率

2.1.2 E ENVIRONMENT 环境比率

2.1.3 M MOTIVATION 动机比率

2.1.4 E EMOTION 情绪比率

refer page 18

**RNN,**

**德塔的词位卷积计算RNN, 主要包含词性比率, 词距比率算子和欧基里德熵算子。这三个算子主要用于求解 POS距离, COVEX距离, EUCLID距离.这些权距 在一篇文章中能够很清楚的计算每一个词汇的使用度，出现的价值，和应用频率以及分布规律。用于文本的主要描述语句的重心所在位置计算。**

1词位卷积计算refer page 178

2用于确定文本的重心

2.1 算子组成

2.1.1 P POS 词性比率

2.1.2 C CORRELATION 词距比率

2.1.3 E E-DISTANCE 欧基里德熵

refer page 18

**DNN,**

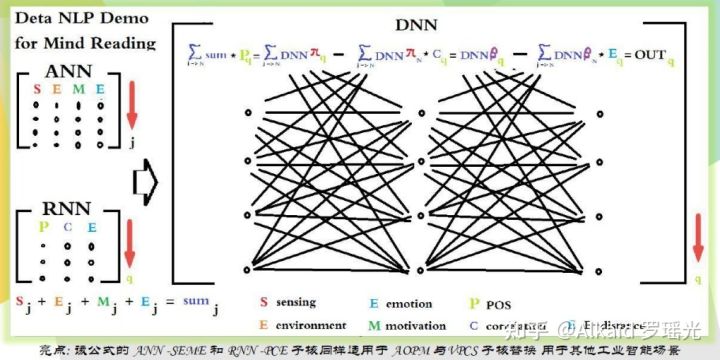
**德塔的词汇深度计算 可以理解为 德塔词性的卷积计算ANN 与 德塔的词位卷积计算RNN 的前序笛卡尔卷积计算。因为参数 由 文章中心思想 和 文章的重心词位 两类组成，因此适用于分析和计算 文章的 核心思想词汇的价值**

1词汇深度计算refer page 183

2用于确定文本的核心

2.1 深度计算 (ANN sum核 -> RNN PCE)

refer page 18



**图灵机，**

1 文学分析refer page 168



德塔文学分析主要用于文章的思想分析和挖掘，如确定多语意识的场景，当时的环境，动机，意识形态倾向和决策思维表达等。（**多语意识 ：通过人物的对话方式，语言特征，模式场景等因素 来 分析当时的人文情感，大众思想，从而了解所处时代的民族风情，社会建筑，时代背景。 教授人：作者导师白育芳，2007年，总参解放军炮兵学院南京分院。**）

**2 作品评估refer page 167**

德塔作品评估 可理解为教育程度评估，如语法，词汇的词性统计，专业词汇的统计，成语，三字词的词长词汇的统计，等等。如一个句子中含有的高级词汇的比率，4字名词的比率，形容词的比率。（作者最早意识出现在**2009年** 在上海章鑫杰那 处理**法国ESIEE亚眠大学**的法语邮件项目， **Pascal教授**曾传授作者关于**FLECH**法语**元音**比重**单词**分析的表述。设计这个项目，进行了灵感发散。**德塔图灵分词全文没有任何单词分析和 非中文的语言分析，不涉及flech任何思想和逻辑，因此一直没有refer**。 作者拥有完整著作权和版权）

**3 动机分析refer page 169**

德塔动机分析 基于动机词典的map key匹配 进行决策表达。比较简单。因为词典定义 带有作者个人主观思维特征。所以没有太多描述。

适用3，4，5

**4 情感分析refer page 159**

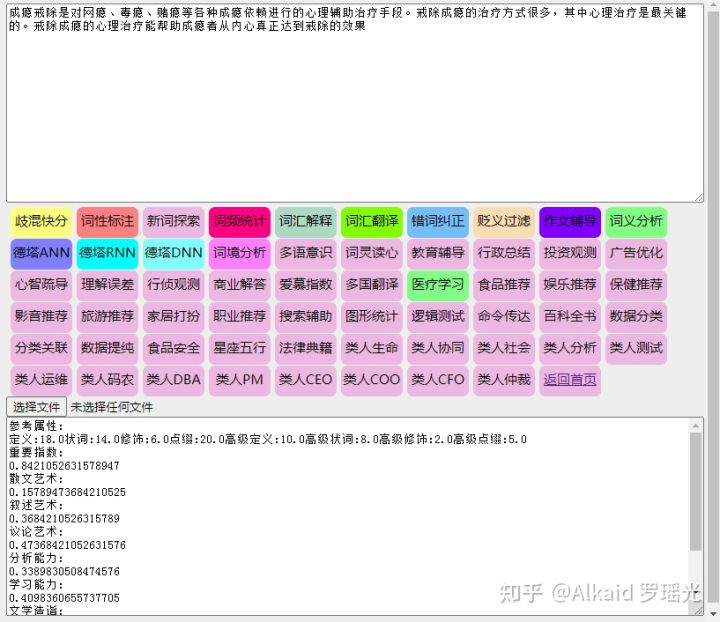
**德塔情感分析 基于 褒义词 贬义词 和中性词 的 map key匹配 进行决策表达。比较简单。因为词典定义 带有作者个人主观思维特征。所以没有太多描述。**

**5 习惯分析refer page 169**

**德塔习惯分析 基于 褒义词 贬义词 和中性词，动机词， 文学分析数据，作品评估比率，教育程度等数据 的全文比重，来确定一个人写作特征，和写作习惯。写作风格。因为词典定义 带有作者个人主观思维特征。所以没有太多描述。**

**6 教育程度评估refer page 168**

德塔教育程度评估体现在文章中的（**有效词汇**如词长超过2位）的 （**有价值词汇**如名动形谓状）的全文，全句，其它POS词性的比率来确定文章的句法特征。举个简单的例子，一个句子中有效有价值的形容词比重大的文章通常代表作者的分析表达和散文修饰能力比较强势。，思维来自作者初中语文学习。



应用

第二章 Java 数据分析算法引擎系统

**定义： 微分催化排序 一般指 将传统的排序在数据排列计算过程中 进行 内存峰值波动平均，计算逻辑减少，计算算子减少，计算条件减少，计算的频率减少，计算关系减少的催化过程。**

价值是方便函数 元基索引 和 新陈代谢，二次新陈代谢。（见之后的 象契排序算法的 新陈代谢催化优化实例）

内存峰值波动平均，(见小高峰过滤左右比对算法)

计算逻辑减少，(见比较函数的 缩进优化)

计算算子减少，(见增序，与减序替换)

计算条件减少，(见狄摩根离散条件or变换)

计算的频率减少，(见选择排序的小于deep的堆栈检测替换，和阀门逻辑序列频率统计 代码排列优化)

计算关系减少 (见算子减少和条件减少的相互关系优化)

的催化过程。

定义人 罗瑶光

函数集合

1 德塔的数据分析包，最早是作者在大学的处理 Rohini教授的 C语言数据结构《Data Structure》 和 Renhart教授 计算机视觉卷积的《Computer Vision》课后作业。

《Data Structure》refer page 226,230,235,238,253,作者没有把当年的计算器四则运算器和rotation tree等作业算法归纳在该作品中。

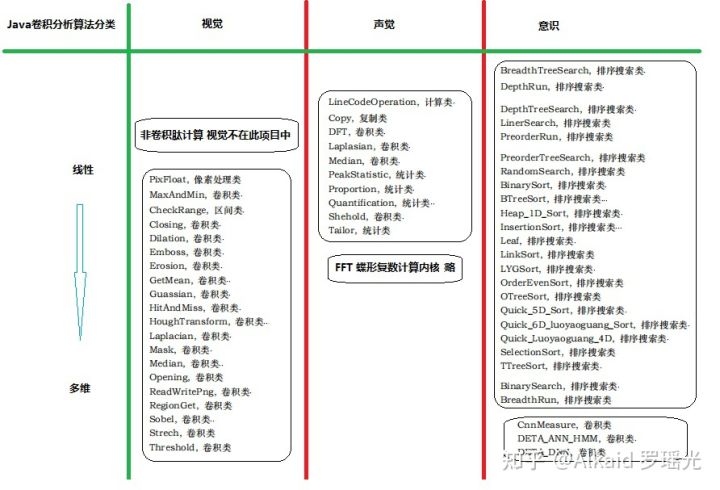
《Computer Vision》refer page 202,204,205,206,209,211,212,213,214,217,218,220,221，259,260，

2 德塔的卷积在2013年后不断的完善，发现其在仿生听觉和视觉计算中都能进行系统的应用，于是开始优化。refer page 191

作者一开始设计卷积是路德大学图片上的应用，2013年，当ETL设计成了节点处理图片像素后，作者开始设计声音 java sound API的处理， 2014年。这个引擎逐渐在计算机仿生系统中进行集成应用。论证了其在具体应用工程中的实践价值。作者当时设计了主要用来测测作者自己的心跳。

3 优化方式为将计算函数进行插件接口模式封装成jar，方便上层调用。refer page 190

4 封装的过程中，不断的进行细化优化，衍生出多个辅助计算函数集，如催化排序，仿生滤波。 refer page 247,655

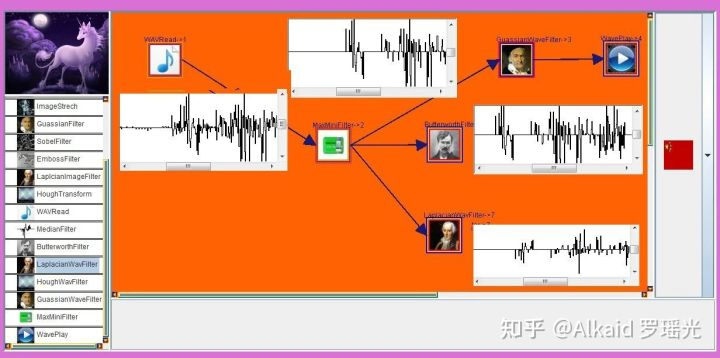
UML

**线性，**

1 德塔的数据分析包 包含array的线性排序处理 refer page 见排序

2 德塔的数据分析包 包含array的线性卷积处理 refer page 见卷积

3 ANN RNN DNN 线性深度卷积计算处理 refer page 222,223,223,



**非线性，**

1 德塔的数据分析包 包含图论的非线性广度建模 refer page 226,230

2 德塔的数据分析包 包含图论的非线性深度建模 refer page 230,232

3 德塔的数据分析包 包含图论的非线性树建模 refer page 236,243,253

**维度，**

1 德塔的数据分析包 包含1维 语音数组计算实例 refer page 见智能声诊

2 德塔的数据分析包 包含2维 图片卷积计算实例 refer page 见智能相诊

3 德塔的数据分析包 包含3维 数据循环阶计算实例 refer page 见噪音识别，三阶傅里叶应用，animation等



**德塔三阶傅里叶计算定义：一般指将线性的时序语音波进行傅里叶变换，此时的波为 频率域波，通过简单的噪声频率过滤后，让后再进行第二次傅里叶变换。于是输出的时序波结果会非常的均匀和格式化，产生优美的平滑间隔峰区间，于是将此时序波第三次傅里叶变换，再次得到的频率波输出具有明确的间隔峰区间生物特征标记。用于德塔语音识别。**

**定义人 罗瑶光**

**场景，**

1 图片的操作。refer page 214

2 像素的操作。refer page 见视觉



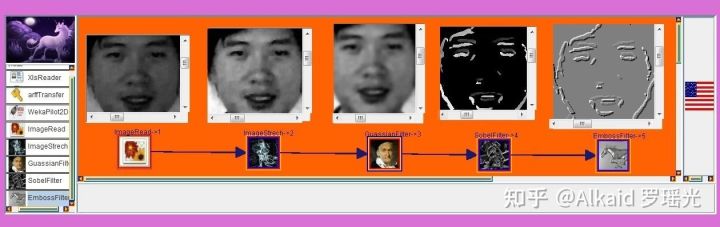
3 文件的存储。refer page 214

4 语音的处理。refer page 见听觉

**仿生听觉，**

1 滤噪计算 高斯1D，median refer page 206, 213, 260

2 频率变换 傅里叶， 快速傅里叶 refer page 258



**视觉，**

1 德塔的视觉主要包含常见2维卷积滤波函数。refer page

2 边缘计算 索贝尔凸蚀，索贝尔梯度，索贝尔向量，拉普拉斯refer page 218, 212

3 凹度计算 emboss浮雕，索贝尔mask，refer page 204,

4 频率计算 傅里叶时序域，傅里叶频率域，哈尔计算，refer page 258, 211,

5 腐蚀计算 膨胀计算，侵蚀计算，均值计算，高斯计算1D一字， 高斯2D十字。refer page 202, 204, 205, 206



**排序，**

1 德塔的排序作者早期2009年设计《算法导论》 黑皮书 ，北邮出版社有其 数据结构 影印教材 的 快速排序4代，进行了10年优化，refer page [https://github.com/yaoguangluo/Data\_Processor/blob/master/DP/sortProcessor/Quick\_4D\_Sort.java](https://link.zhihu.com/?target=https://github.com/yaoguangluo/Data_Processor/blob/master/DP/sortProcessor/Quick_4D_Sort.java" \t "https://zhuanlan.zhihu.com/p/_blank)

2 优化过程归纳，逐渐的形成了一个微分催化排序体系。refer page 247,248,250, 658，下册134，

3 左右比对算法优化，小高峰过滤优化，缺陷峰归纳，催化算子优化，离散逻辑优化。refer page 658，下册134，

**左右比对算法优化，**一般指在不对称的数列中，为了寻找对称性观测面，作者设计了一种比较简单的方法，如将数列逐渐拆分， 取出拆分后的小数列的初值和尾值进行比较，作为一个参照点，用于躲避计算高峰。测试发现具有强大的实用性。

**小高峰过滤优化，**一般指为了躲避内存计算高峰而导致的延迟，卡顿，死锁，堆栈溢出等问题 而设计的一类高效率算法集合。

**缺陷峰归纳，**一般指计算数列在不断的拆分中的中值基偶问题导致了变量，算子，函数的使用频率不对称而出现的一系列蝴蝶效应问题集的归纳。

**催化算子优化，**一般指 计算的中间过程中 因 变量，算子，函数的使用频率 不对称，不稳定导致的各种问题 ，为了解决这类问题而 进行的不断 的对 变量，算子，函数优化与校正过程。

**离散逻辑优化，**一般指 对 变量，算子，函数优化与校正过程中 通过离散数学， 迪摩根定律，等客观存在的逻辑定律进行 不断优化与校正过程。

**定义人 罗瑶光**

4 目前代表作为TopSort5D 极速催化排序。refer page 下册134

**搜索，**

1 德塔的搜索计算主要做一个编码参照，没有工程用途。refer page 226

2 编码参照有利于作者在设计图论计算和非线性搜索时候发散思维用途。refer page

3 编码开始于作者2009年 完成 Rohini教授布置的作业。refer page 我 qq 313699483 有完整作业备份日记。

4 对作者研究Hash空间 有发散思维的用途，如作者数据预测包设计的辅助。refer page

**应用**

1 TopSort5D 包含深度算子，包含广度算子，包含滤波算子。refer page 下册134

TopSort5D 版权源码

本人调通的算法导论的quicksort4D算法链接如下，可直接区别，再次Refer 快速排序之父 霍尔先生：

[https://github.com/yaoguangluo/Data\_Processor/blob/master/DP/sortProcessor/Quick\_4D\_Sort.java](https://link.zhihu.com/?target=https://github.com/yaoguangluo/Data_Processor/blob/master/DP/sortProcessor/Quick_4D_Sort.java" \t "https://zhuanlan.zhihu.com/p/_blank)

2 索贝尔 dir 向量差 区别三维的立体面特征趋势。refer page 219

3 噪音识别。refer page 720

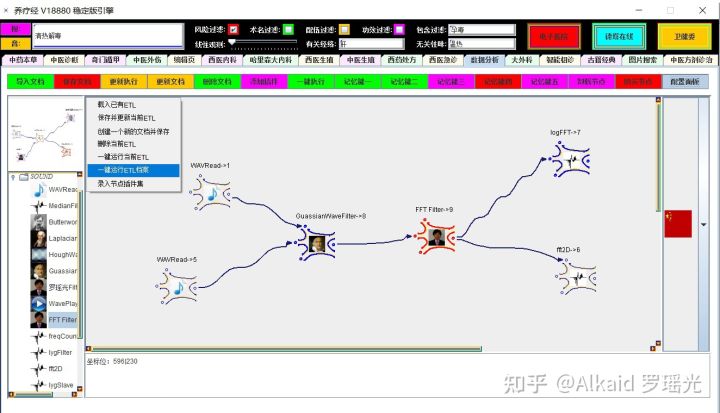
4 小波分离。refer page 不在此章 涉及鸡尾酒调度，被略去先

5 极速商旅TSP。refer page 538,541,547

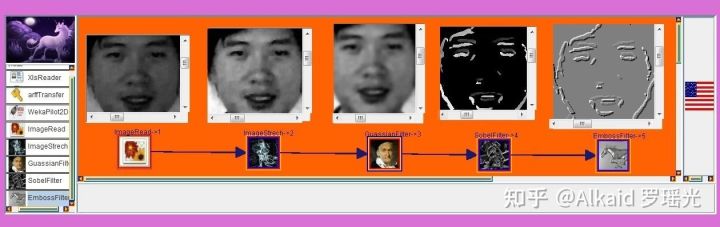
6 股票数据抓取 refer page 不在此章， 261, 263, 264, 266可以处理 股票数据线波。

**涉及著作权文件：**

第三章 德塔 ETL 人工智能可视化数据流分析引擎系统.



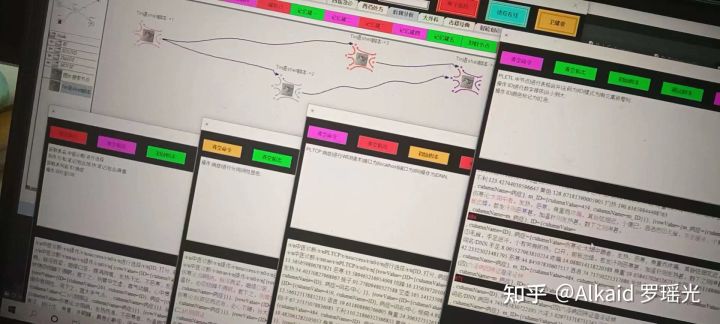
1 德塔ETL 又叫 ETL UNICRON, 是一个数据节点流计算的可视化操作工具.refer page 267



2 最早由作者在路德大学设计java卷积视觉包，为了方便 像素矩阵流的流水观测。refer page 186

3 德塔ETL采用APPLET,可以嵌入在网页上作为 rich web架构，与flash应用相似。refer page 287

4 德塔ETL已成为当前的PLETL,和元基花模拟神经元计算的基础组件。refer page 774



**界面，**

1 德塔ETL 的界面采用Jsplitpane分区 主要包含节点显示树区，节点画布操作区，计算状态反馈区 和系统配置区。refer page 286~

2 节点的显示区 采用Jtree进行鼠标操作，左键选择，右键弹框。refer page 286~

3 画布操作区 的 节点采用画线描点实现，操作为左键拖拽，右键连线和弹框。refer page 311~

早期的节点处理界面弹框 设计成inner弹框模式作者发现关闭按钮被屏蔽了，于是就改成frame组件跳出canvas画布来显示节点处理界面。

4 状态反馈与系统配置区用于实时了解ETL的运行状态。refer page 见jtextPanel

**最新 养疗经 版本，作者将状态数据的jtextpanel 写在了元基枝全局变量进行集成，如果要单独将etl做插件使用，要区别这个 函数，可以改写下。**

**皮肤，**

1 德塔ETL 皮肤采用bitmap实现，可以自由替换。refer page 333,334

（**之后我会进行代码的优化优化， 将 bitmap的表达形式 array[][] swap成 array[]+array[]，这样50\*50的界面 含有200个点， 2500次遍历 就变成了200次， 界面刷新时间缩减到原来的十分之一甚至更短。罗瑶光**）

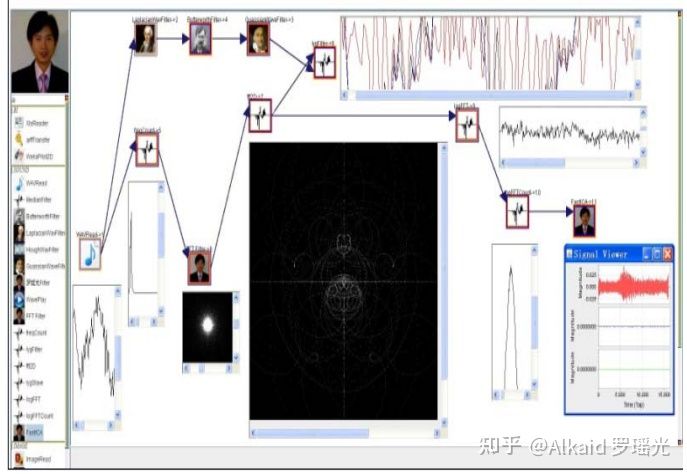
**写清楚点， array[10行][10位]= 1 or true 如果坐标矩阵第10行第10位 是像素显示标记1， 那么可以swap成 array[0]=10行， + array[0]=10位， 这样极大的减少计算array的内存buffer占用。再举例 array[13行][9位]= 1 or true 如果坐标矩阵第13行第9位 是像素显示标记1， 那么可以接下来swap成 array[1]=13行， + array[1]=9位，方便理解。**

2 德塔ETL 皮肤包含指标，箭头，连线，节点外形，控间外形设计。refer page 332,319~

3 德塔ETL 的控件采用jdk的组件 componient进行2次开发。refer page 334~

4 德塔ETL 的引擎界面的描点画线基于AWT Swing的canvas 画布系统实现。refer page 311~

另外德塔早期曾用界面 LYG-AI 如下：



**流存储，**

1 德塔ETL的流存储是一个object类，可以包含多种状态。refer page 329~332

2 德塔ETL的流采用单例的this.clone,确定了中间态存储模式。refer page 见 节点 clone()

3 德塔ETL的流可保存，可观测，但不可逆。refer page 329,330

**节点，**

1 德塔ETL的节点作为一个计算单元，模拟神经元作为最小单位计算。refer page 327~

2 德塔ETL的节点存在UI多样化，可界面设计如orange， knime，weka那样。refer page 335

3 德塔ETL的节点也可以语言化，如PLETL的语句IO模式研发设计。refer page 774~,790

4 德塔ETL的节点大体为从左到右的IO模式，节点流计算目前可循环但不可逆。refer page 329~332

**[德塔可视化ETL操作例子](https://www.zhihu.com/zvideo/1483183087273684992" \t "https://zhuanlan.zhihu.com/p/_blank)**

[](https://www.zhihu.com/zvideo/1483183087273684992" \t "https://zhuanlan.zhihu.com/p/_blank)

[Alkaid 罗瑶光的视频](https://www.zhihu.com/zvideo/1483183087273684992" \t "https://zhuanlan.zhihu.com/p/_blank)

[· 11 播放](https://www.zhihu.com/zvideo/1483183087273684992" \t "https://zhuanlan.zhihu.com/p/_blank)

**插件，**

1 德塔ETL的插件类似OSGI的jar开发模式。严谨的说只是继承的classloader模式。refer page 286,290,777

2 德塔ETL的插件可以加元基标识认证组件，避免错误插件扩展。refer page 290,777,779

3 德塔ETL目前支持插件进行平台配置，页面扩展和节点扩展。refer page 286,290

4 德塔ETL目前的3元基文件名索引肽化支持插件的分类管理和加密标识。refer page 781

**档案，**

1 德塔ETL的档案包含节点流信息和节点配置信息。refer page 279,282

2 德塔ETL的存储采用节点的画布状态单例信息存储方式.refer page,279,,282

3 单例信息包含画布中节点的坐标，名称，ID，连线，配置信息等实体信息。refer page 282

4 德塔ETL的流存储用文件读写形式.etl后缀存储.refer page 282

5 德塔ETL的存储可支持加密和batch模式运行。refer page 养疗经应用略

**拓扑，**

1 德塔ETL的拓扑体现在节点的神经元模拟计算观测。refer page 273

2 节点的神经元模拟拓扑体现在从左到右的从高到低拓扑模式。refer page 273

3 第2点或许是个拓扑缺陷，但是却因此又确定了固有 的向量方位。refer page 273

**神经网络，**

1 德塔ETL的神经网络计算在流数据计算中有实际价值。refer page 274

2 德塔ETL神经网络在PLETL和tinshell中逐渐体现其更多学术价值。refer page 783

3 德塔ETL的向量拓扑模式确定了神经网络的加权方式。refer page 274

4 德塔ETL的神经网络可循环不可逆。refer page 274

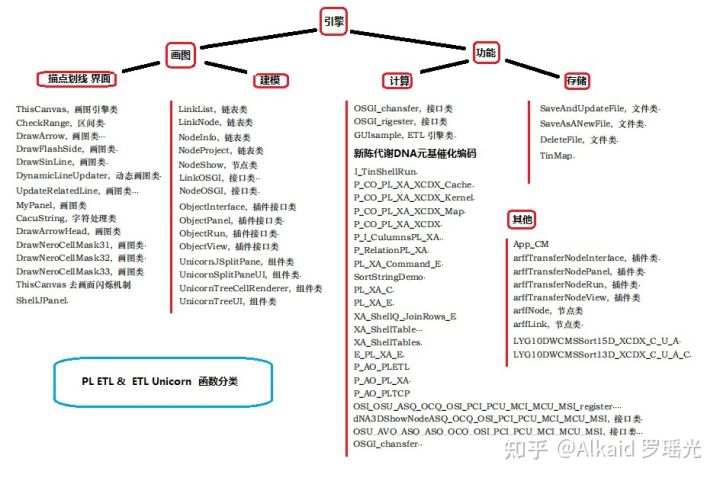
**一键执行，**

1 德塔ETL支持 一键保存 。refer page 277, 282

2 德塔ETL支持 一键读取。refer page 302

3 德塔ETL支持 一键执行。refer page 308

德塔的一键执行最早作者想设计成Knime的那种batch的批处理脚本节点流那样，随着作者TinShell出来后，作者觉得批处理跨语言，即繁琐又没效率。于是想做成JAVA脚本语言，伴随着个思路，目前作者设计了TinShell语言和元基花语言。



第四章 德塔 Socket 流可编程数据库语言引擎系统

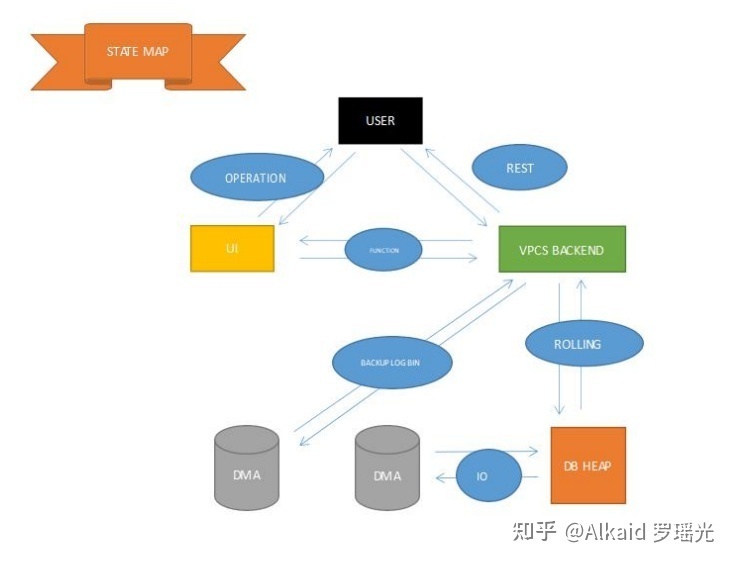
**Socket rest TCP握手协议，**

1 德塔数据库的 admin界面采用 web页进行配置操作。refer page 376

2 web页配置操作采用TCP握手访问模式，基于socket的http请求握手。refer page 464~

3 德塔数据库将socket握手进行线程封装，然后多线程组织页面。refer page 392，

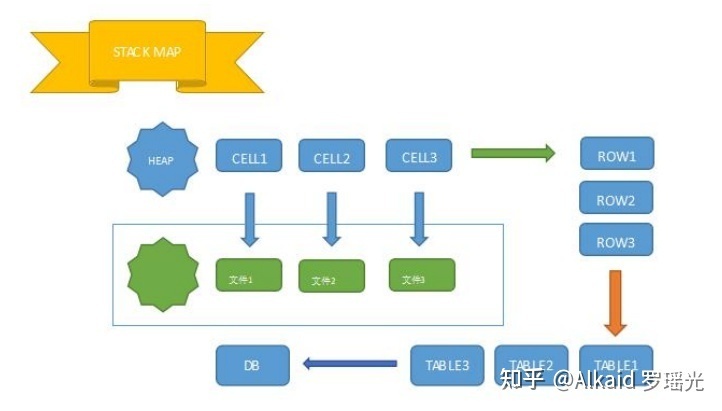
4 封装和组织页面设计过程逐步进行优化形成VPCS后端管理体系。refer page 383,476



**文件数据库，**

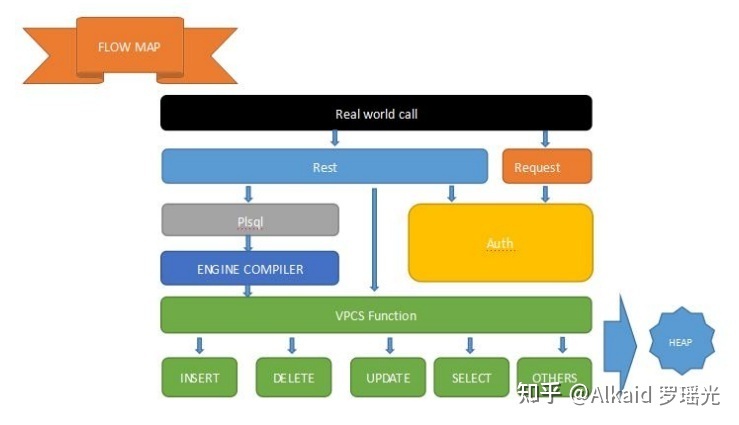
1 德塔数据库的数据存储是一种文件存储模式。refer page 408,409,469,473

2 文件的读写进行子集，行，表，映射，表头，按1范式分类。refer page 375,434,



3 数据库的数据读写支持加密。refer page 见元基加密

4 每一个文件不但有物理空间，还有相应的内存空间。refer page 375



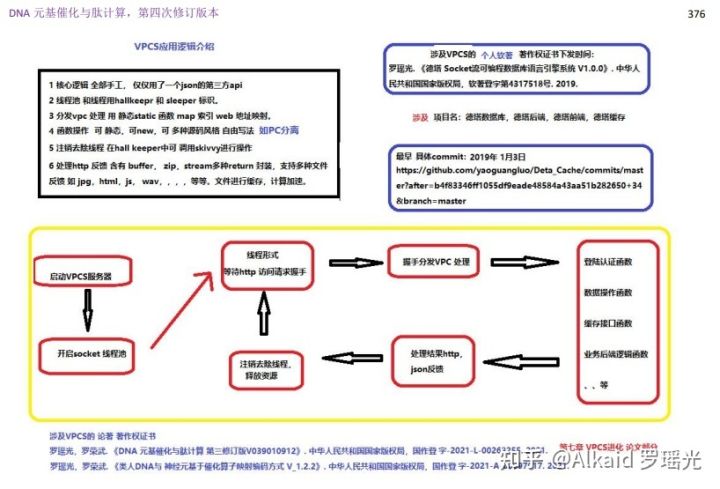
**VPCS服务器，**

1 VPCS服务器支持每秒400万QPS的web请求。refer page 389,

2 VPCS服务器采用TCP rest request模式，标准化http response。refer page 388,395

3 VPCS服务器可自由设计前端和后端集成。refer page 见德塔官网 和 养料经admin 两个实例

4 VPCS服务器完全支持post 个 get 2种请求模式，可扩展。refer page 481,488



**VPCS调度架构，**

1 VPCS服务器包含 视觉模块，处理模块，控制模块，资源模块。refer page 396,394,392,383

2 每一种模块有各自的名称标识 和 内存标识，方便精确查找。refer page 492,493,

3 VPCS服务器包含执行者-生产者-造梦者-sleeper，管理者-分配者-登记者-HallKeeper，运维者-服务员-清洁员-skivvy 3个模式。refer page 392,394,

4 支持控制与执行分离，线程与资源分离。refer page 385~389,486,490,492

**作者最早设计 VPCS 服务器的动机，是为了弥补VPC的计算过程观测困难的问题。因为作者设计的VPC是采用springboot + mybatis的结构，底层全是是开源插件的封装，很多核心源码又不能调试仅仅通过几个log和 trycatch给作者带来了无形的压力（作者的思维很简单，就是自己写个服务器，能够调试断点从头断到尾），于是有计划从无到有进行设计一个TCP/IP的 SOCKET 协议做服务器HTTP请求。作者当时没有想到，一个这样的小动机给带来了丰厚的回报，VPCS 目前成为了 DNA 元基映射编码算子 的核心组成部分。**

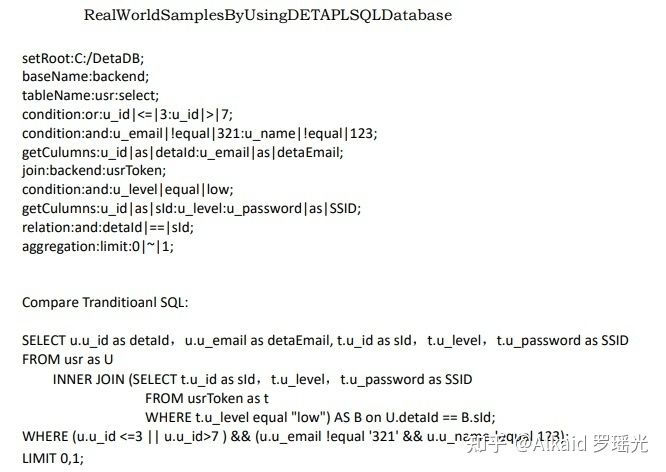
**PLSQL语言，**

1 德塔PLSQL语言是一种从上到下的脚本执行语言。refer page 377,

2 德塔PLSQL语言包含常用增删改查命令。refer page 406~409,471,1035

3 德塔PLSQL语言支持join和 aggregation 高级操作。refer page 419,431,435,438,447

4 德塔PLSQL语言行 可批处理，可拆分。refer page 1035~1041 将例子写入main，class编译，然后 bash boot class 即可。 还可以bash 定时批处理。



**PLSQL编译机，**

1 德塔PLSQL编译机 用于理解和执行 德塔PLSQL语言。refer page 413,414

2 德塔PLSQL编译机 包含常见脚本命令计算算子如 条件算子，比较算子，包含算子，离散算子。refer page 419

3 德塔PLSQL编译机 采用map进行的内部中间数据缓存。refer page 431,432~

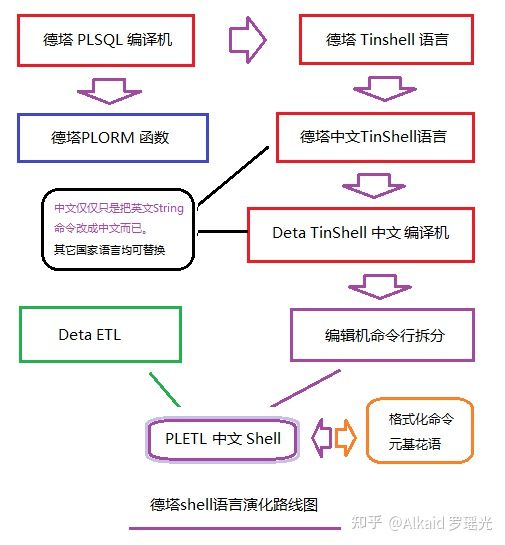
**PLORM语言，**

1 德塔PLORM语言 用于 德塔PLSQL语言进行函数封装。refer page 1003~

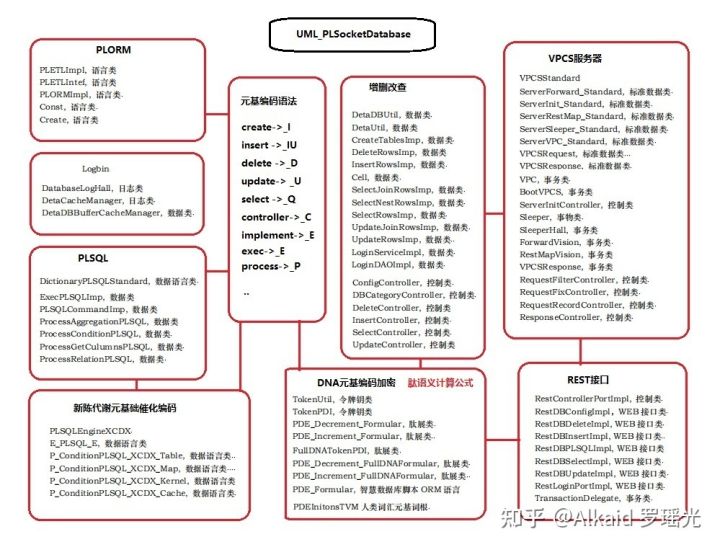
2 德塔PLORM语言 有先后顺序，需要遵循 德塔PLSQL语言语法。refer page 1019~

3 德塔PLORM语言 对比 德塔PLSQL语言 用于一些不需要配置的nosql的场景，类似 hibernate 对比 ibatis。refer page 1019~

4 VS hibernate 对比 ibatis的不同，德塔PLORM语言 另外也是 德塔PLSQL的上层语言。refer page 1019~



**德塔的PLORM 和 PLSQL 的引擎出现，作者开始有信心将其优化成 节点执行的命令行脚本模式，于是之后的TinShell 和 PLTinShell， PLETL Shell 诞生了。这个PLETL体系弥补了 当前世界按语言理解方式来模拟神经组织计算的映射空白。**



**灾后重建，**

1 德塔数据库包含logbin 系统。refer page 398,

2 德塔数据库包含logbin 系统基于单个写操作进行log保存 并行加密成文件。refer page 399

3 单个写操作用时间戳作和写增量序列进行对应标识，避免混乱。refer page 399

4 德塔数据库包含logbin 系统 并支持热备和错误写 实时rollback 检测。 refer page 398

**第五章\_德塔数据结构变量快速转换**

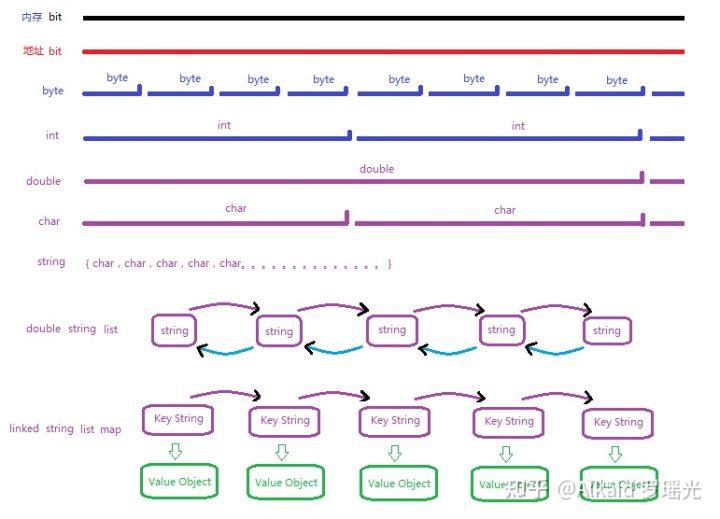
**内存的结构，**

1 德塔数据结构变换 最早归纳来自对 雪球新浪的股票数据 web页抓取进行的String格式统一。refer page 508,528

2 基于String的格式统一，然后逐步进行文本数据在计算过程中的状态进行分类扩展归纳。refer page 532,535

3 于是产生array，StringBuilder， iterator，map， 4种 高频内存结构的快速互换。refer page 499,536,515,520

4 最后进行对所有常见数据结构进行统一归纳和快速变换。 作者的研发基础来自2008年 在印度基督大学的C语言数据结构实验室课程。讲课教授 Rohini.V refer page 492~

常见数据结构类型，罗瑶光画图

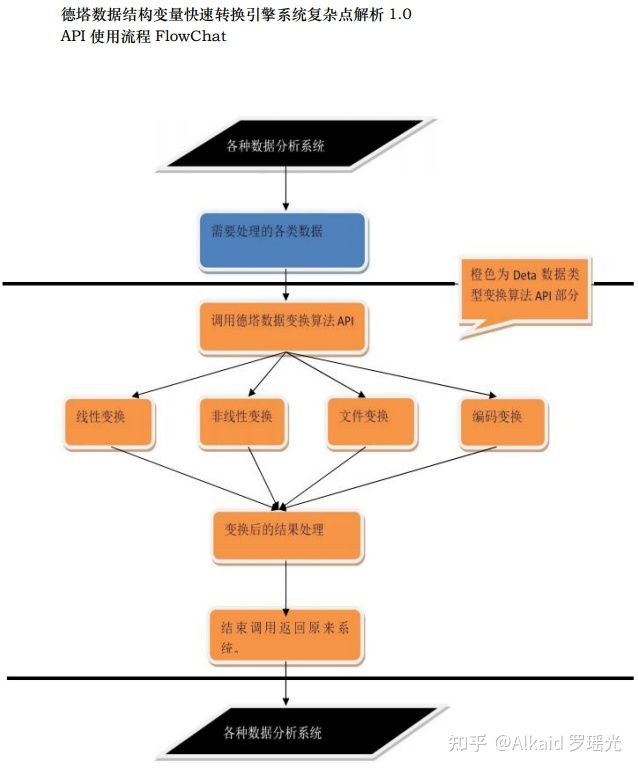
**数据的结构，**

1 德塔数据结构梳理完整依据 C语言数据结构 思维进行归纳refer page 无

2 归纳包含 array，String，struct object，hash， map list， tree，buffer的数据变换模式。refer page 499，535,527,507,520,516,537，

3 德塔数据结构不包含数据的计算逻辑变换，仅仅包含数据类型的载体变换。refer page 498

4 数据类型的载体变换通过接口形式表达。广泛用于工程中。refer page 498



罗瑶光画图

**类的结构，**

1 德塔数据结构的类，采用VPCS的静态接口模式设计。refer page 492~

2 每一种相同数据类函数封装在同类的文件中。refer page 492~

3 每一个类 主要包含数据变换文件，数据变换的纠正文件，数据变换的索引文件。refer page 492~

**转换加速，**

1 数据变换的索引文件，通过元基花索引24组染色体注册，进行语言调用加速。refer page 下册597 StaticFunctionMapU\_VECS\_E

2 数据变换采用静态函数，加速了function call。refer page 492~全章

3 数据变换的函数 根据功能进行了分类，于是静态函数文件形成了balanced静态函数集树模式。refer page 下册274 第十六章

**不规则对象的变换，**

1 不规则对象的变换主要包含 邻接矩阵array变换和 类复制。refer page 521

2 邻接矩阵array变换 如 跨格式变换，如xml，json，officerefer page 558,516,503

3 类复制如 deta的TinMap class和 Objectrefer page 527,881

4 xml和json，德塔不做加工，仅仅用google的Gson包引用。refer page 516

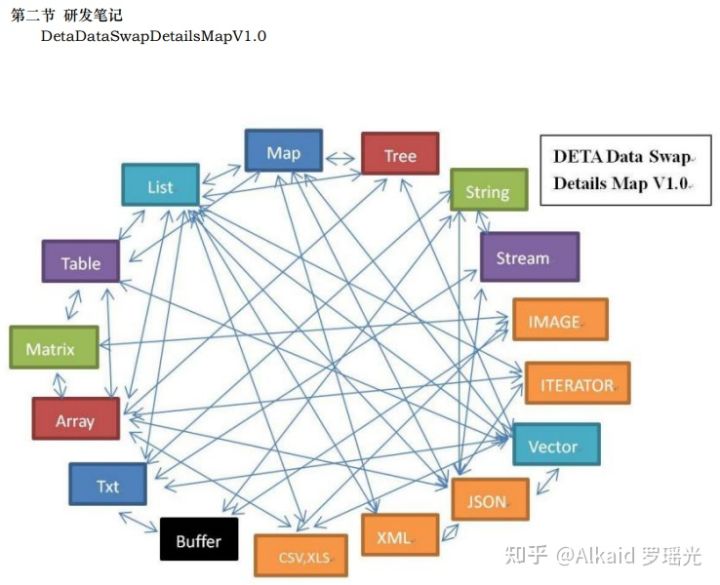
**场景变换，**

1 德塔数据结构的场景主要应用在网页html数据抓取，文本数据计算refer page 508,492~

2 html数据 主要体现在文字的编码格式变换，加密变换，和http response的内容载体变换如json。refer page 508,555，

3 文本数据计算主要体现在 map和array的变换，与 list和array的变换，用于字符串排序加速。refer page,499,516,520

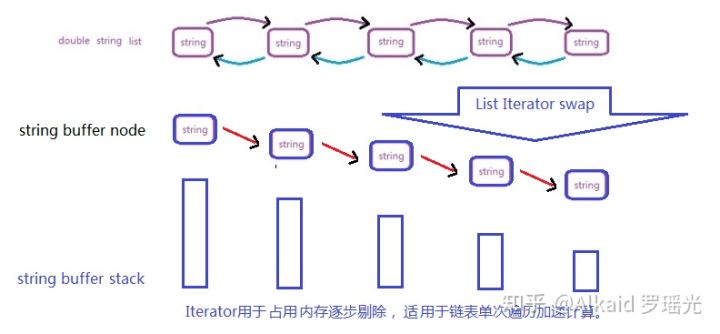
4 在德塔分词场景中体现在另外String与 StringBuilder的加速变换。refer page 536

罗瑶光画图

**计算的模式变换，**

1 德塔数据结构计算的模式变换主要用于 buffer中间态变换。refer page

2 buffer中间态包含 map与tree的变换， list与iterator的变换。refer page 520,537,516,515



图中String buffer stack 可先后序列排列，可断开成链，高度是iterator对象当前的内存占用大小。罗瑶光画图

list.toIterator()变换模式优势罗瑶光先生个人认为在计算过程中，基于内存的占用和寻址效率加速。String to Stringbuilder 变换同理，对象buffer化能实现内存变量计算和调用进行极限加速。作者在印度基督大学 学数据结构没有stringbuilder和 iterator的知识点，在2016年亚马逊的岗位技术经理面试时候， 有几次印度经理多次面试我关于String计算方式，我当时没有答上细节，错失了月薪12000美金的工作。我的罗瑶光画图

3 模式变换计算趋势归纳主要为非线性与线性的降维变换，通过改变观测面实现。refer page 497

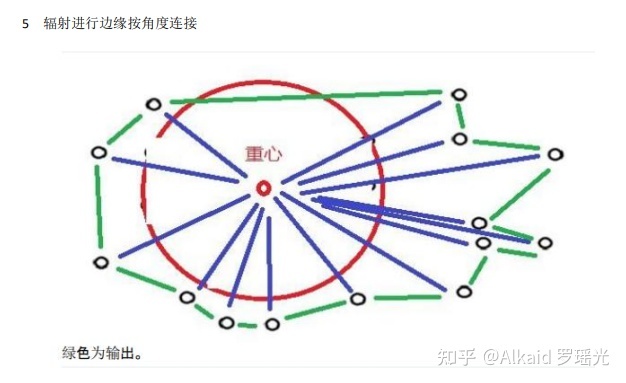
应用

罗瑶光画图

太多略。

**章节的著作权文件列表：**

第六章\_数据预测引擎系统



此算法为版权算法

api包 函数完整包含2维和3维的空间轨迹算法。

[GitHub - yaoguangluo/Data\_Prediction: 快速计算商旅轨迹 非线性坐标数据分析](https://link.zhihu.com/?target=https://github.com/yaoguangluo/Data_Prediction" \t "https://zhuanlan.zhihu.com/p/_blank)

Java api [https://github.com/yaoguangluo/ChromosomeDNA/blob/main/BloomChromosome\_V19001\_20220108.jar](https://link.zhihu.com/?target=https://github.com/yaoguangluo/ChromosomeDNA/blob/main/BloomChromosome_V19001_20220108.jar" \t "https://zhuanlan.zhihu.com/p/_blank)

**坐标系统预测，**

1 数据预测引擎的坐标系统主要用来做离散非线性计算。refer page 566~

2 离散非线性计算主要体现在 降维 商旅TSP路径的线性求解。refer page 629~

3 坐标的降维计算包含 轨迹降维，趋势降维，观测降维。refer page 567~

4 降维计算过程可以进行逆向跟踪还原。refer page 570

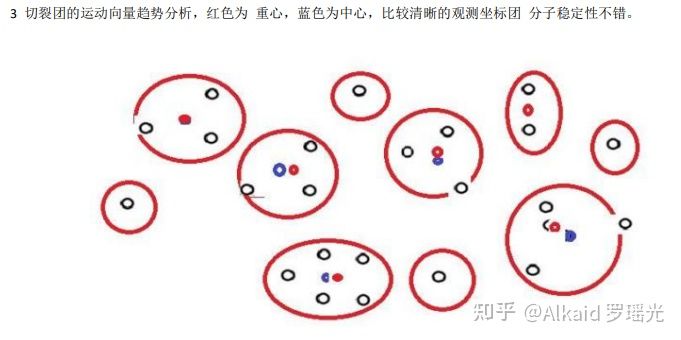
**环境预测，**

1 数据预测引擎的环境计算主要体现在 压力计算。refer page 570,573

2 压力计算可理解为 中心向重心的两点间距离。refer page 674

3 两点间距离的长短和方向代表压力的大小和趋势。refer page 574 红蓝点距离

4 趋势大小确定环境的稳定性表达。refer page 574



**雷达机，**

1 数据预测引擎的雷达机主要体现在坐标群的边缘识别和归纳计算。refer page 577

2 坐标群的边缘识别和归纳计算 采用角度 + 中心到点距离进行进行轮循链接。refer page 576

3 链接的面形成 极速计算边缘包含，确定坐标的团大小面积，密度。refer page 577

4 极速边缘计算的价值可以迅速利用在所有实时坐标系统中。refer page 593

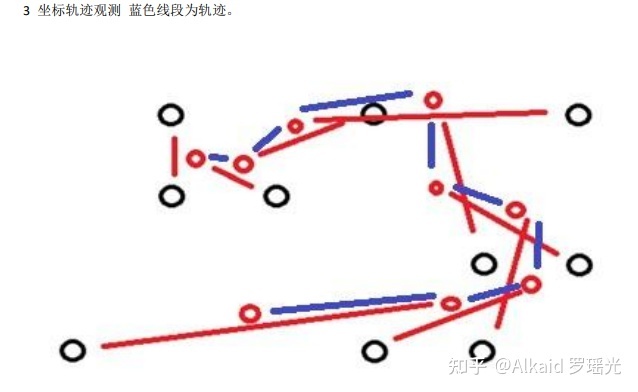
**状态机，**

1 数据预测引擎的状态机主要包含 压力状态，轨迹状态，refer page 571，573，

2 压力状态体现在坐标团的之间的距离，和团中心和重心的距离分析。refer page 571

3 轨迹状态体现在坐标团的内部欧基里德距离熵增和团中心KNN迁移熵增分析。refer page 569,570

4 数据预测引擎的状态机应用在非线性坐标计算系统中。refer page



**离散模型预测，**

1 数据预测引擎的离散模型预测，作者主要用在商旅计算中。refer page

2 作者主要用在商旅计算中的 小坐标分子群计算中。refer page 568

3 作者的商旅计算最大价值主要体现在 欧拉环路的分析中。refer page 568

4 作者的 欧拉环路为破解 十六进制 十六元基进制编码 起到了基础研究作用。refer page 下册56, 下册125

**概率机，**

1 数据预测引擎的概率机比较简单，仅仅贝叶斯系统。refer page

2 贝叶斯系统在作者的工程中很少用到，如线性回归，衰变失效就不包括。refer page

3 贝叶斯系统作者有设计交叉概率机，关于数据挖掘pangningtan教材的质量分析。讲课教授 卡拉森。refer page 616

4 作者设计概率机，主要是之后做图片识别预测用。refer page

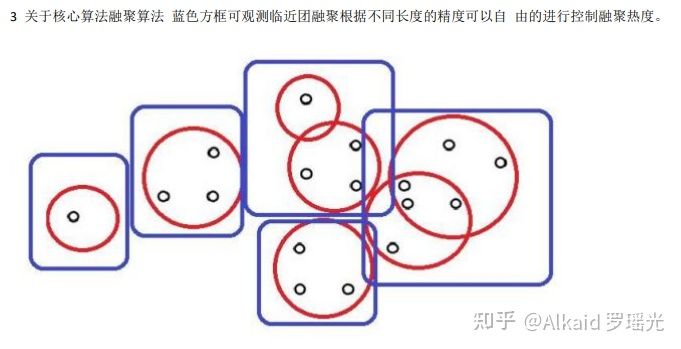
**向量机，**

1 数据预测引擎的向量机作者主要设计了团中心和重心的距离向量。refer page 595

2 距离向量 可以作为路径猜测，运动趋势，和轨迹判断用途。refer page 621,624,634

3 距离向量理解为斥力，可以表达坐标团的稳定性评估。refer page 601

4 距离向量理解为压力，与雷达机结合，可以计算表达坐标团的密度。refer page 610,613,605, 593



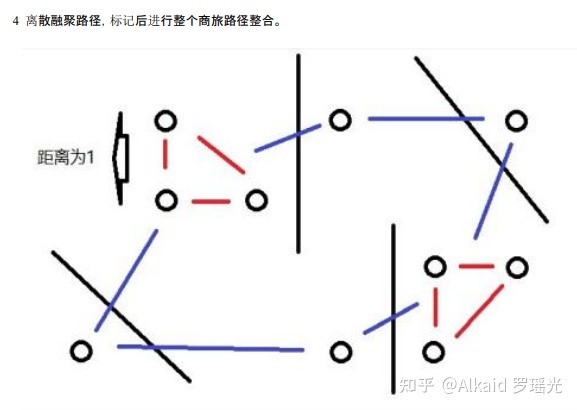
**德塔坐标团的密度 一般指，将坐标进行 观测距离的区间进行划分后的坐标融聚小团，的坐标数和团数的比值举例 如果划分有5个区间，每个区间坐标数是 1,3,4,3,6,，那么比值是1/5 , 3/5, 4/5, 3/5, 6/5 这里的观测距离是可以精度调节的。通过排序可以迅速计算 用于确定压力的位置。 定义归纳人 罗瑶光，稍后优化**

**商旅TSP，**

1 数据预测引擎的商旅TSP，主要计算随机坐标集的欧拉环路。refer page 625

2 数据预测引擎的商旅TSP，作者设计动机为极速小分子团间的欧拉2阶图研究。refer page 630

3 作者研究动机为破解元基罗盘的 离散活性邻接矩阵变换。refer page 下册5，



4 作者研究结果为十六元基进制 破解 DCPE-THOS-MAXF-VIUQ 。refer page 下册5，下册56,下册125

应用

太多了，略，

**早期应用实例，不仅在德塔自己的坐标插件可以灵活应用， detaETL 也可以集成 awt+ weka第三方插件研发 进行数据显示实现。如下图的pilot例子。作者早期用swt+knime进行weka设计，自从自己写了etl unicorn后，发现SWT插件都不需要了。**



另外函数分类方法如 切裂，融聚，隔离，簇类，就不介绍了数据挖掘的聚类思想作者个人表达方法而已。

