

# 论文写作

# 张心怡

2025.03.27

### 单元内容

- ▶论文写作为何重要
- ≻写作树
- ▶好的写作风格
  - 精炼
  - 正确的语气与态度
  - 例子
  - 动机
  - etc.
- ▶具体风格
  - 标题与小标题
  - 开头段落
  - 段落
  - etc.

### 单元内容

- ▶论文写作为何重要
- ▶写作树
- ▶好的写作风格
  - 精炼
  - 正确的语气与态度
  - 例子
  - 动机
  - etc.
- ▶具体风格
  - 标题与小标题
  - 开头段落
  - 段落
  - etc.

### 论文写作为何重要

- > 科学论文的核心任务:
  - 提出新观点
  - 论证其正确性
- ▶论文具有长期的学术价值。
  - 好的写作能提升研究的影响力,差的写作可能导致 成果被忽视。

### 论文写作为何重要

- ▶糟糕的写作会导致理解困难,降低研究的说服力。
- 模糊不清容易引发误解。
- 表达不清会浪费读者时间,削弱研究效果。
- 再好的故事也能被糟糕的讲述方式毁掉。
- >写作不是附加技能,而是科研中的核心技能之一。

### 论文写作为何重要

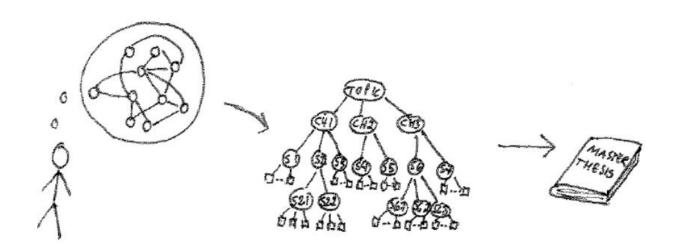
- >写作并不是研究的终点,而是研究的一部分。
  - 写作促使你澄清模糊的概念、构建严谨的论证。
  - 写作是将思维结构化的重要方式。
  - "写清楚"往往也是"想清楚"的过程。
- ▶科学知识必须"写出来"才成立。

### 论文写作

•写作 w是从一组想法 I到一组科学文本 S 的映射:

w: I→S

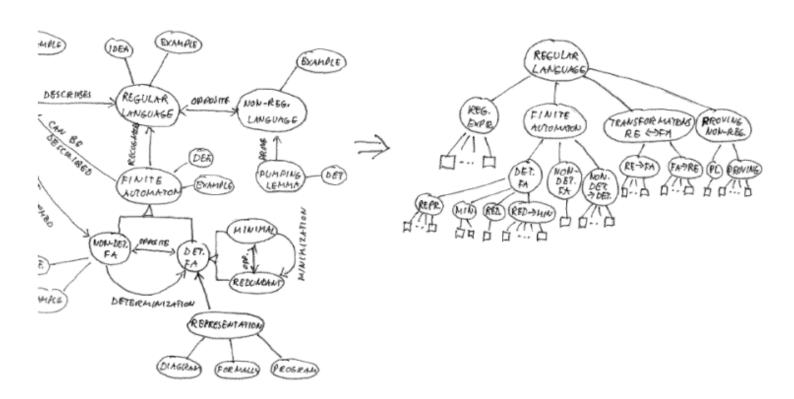
• 问题: 给定一组想法 i∈I, 生成 f(i)∈S。



### 想法树

➤以树状结构 (想法树 t) 组织你的写作思路。可理解为

:在"想法图"上构建"最小生成树"。



#### 想法树

- ▶将这棵树 t 以文本形式写出, 要求如下:
  - 根节点代表主题 (标题)
  - 根的子节点代表章节 (chapters)
  - · 子节点的子节点或孙节点代表小节 (sections/subsections)
  - 叶子节点代表正文段落(实际文本)

#### 一棵好写作树的特征

- ☑ 结构平衡
- 所有从根到叶的路径长度大致相同
- 通常 ≤ 4, 最多不超过 5 层
- ☑ 每个节点的子节点数合理
- <del>一</del>般 k ≤ 7, 上限为 k = 10
- 最理想的是 k ≤ 2 (结构清晰)

#### 一棵好写作树的特征

- ☑ 段落长度平衡
- 所有叶子节点 n(段落):
  - content(n) 至少包含两句话
  - 不应太长(推荐≤7~10句)
- 所有非叶子节点 m (章节或小节):
  - content(m) 为简介,可非常简洁
  - 简单介绍该部分要讲什么

#### 一棵好写作树的特征

- ☑ 正文引用规则 (按先后顺序)
- 只能引用前面已写的内容 content(n₁), ..., content(ni-1)
- 💢 不允许 "倒引用"

例如:不能用"确定性自动机是非确定性自动机的对立面",而前面尚未定义"非确定性自动机"

- 🗹 允许"预告未来内容"
- 如: "该问题将在第 X 章中解决。"

#### 单元内容

- ▶论文写作为何重要
- ▶好的写作风格
  - 精炼
  - 正确的语气与态度
  - 例子
  - 动机
  - 平衡性
  - 语态

### 写作风格

- >风格不是语法,而是你用什么方式和读者交流思想。
  - 英语表达可以繁复或简洁、华丽或朴素。
  - 写作风格决定读者的理解难度和阅读体验。
  - 科学写作应追求清晰、准确、直白,但不代表必须 枯燥无味。

#### 风格误区

• 故意炫技: 堆砌复杂词句, 让人难以理解。

• 过度修饰:用文学语言掩盖科学内容。

• 违背规范:任意打破惯例,导致注意力被分散。

• "吸引眼球"≠有效沟通

→每个句子都应有其存在的必要性,删去冗余文字,提 升可读性。

- · 检查是否有"废话"(waffle)或"填充句"。
- 每句话都应推动内容前进。
- 多次修订,去除冗余

#### • 举例: 例子中的斜体部分可删而不影响含义

The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database.

- ▶简洁写作的原则
  - 1. 保持结构简单且逻辑清晰。
  - 2. 使用简洁的词语。
  - 3. 使用结构简单的短句。
  - 4. 保持段落简短。
  - 5. 避免使用流行语、陈词滥调和俚语。

- ▶简洁写作的原则
  - 6. 避免冗长,无论是篇幅还是风格。
  - 7. 删除所有不必要的内容。
  - 8. 表达要具体,不要含糊或抽象。
  - 9. 如有正当理由,可以打破以上规则。

#### 正确的语气与态度(Tone)

• 警惕"过度谨慎"的语气

•写作≠推销,不要文学化表达

• 不要模仿"科普文风"

#### 警惕"过度谨慎"的语气

- 不要用大量限制词"埋掉"你的主张。
- 科学写作不是回避判断,而是基于证据地做判断。
- X The results show that, for the given data, less memory is likely to be required by the new structure, depending on the magnitude of the numbers to be stored and the access pattern.
- ✓ The results show that less memory was required by the new structure. Whether this result holds for other data sets will depend on the magnitude of the numbers and the access pattern, but we expect that the new structure will usually require less memory than the old.

#### 不要文学化表达,不要模仿"科普文风"

• 学术写作要准确、清晰, 避免"拟人化""命名化"

修饰语。

#### 不推荐:

- the "heart" of the system
   the "Grimwade effect"
- ・推荐表达:
- the central server
- We observed this effect during...

### 用例子让抽象变得具体(Example)

• 一个例子解释一个概念,例子比定义更易理解

- 简短例子胜过长篇讲解
  - ✓ In a semi-static model, each symbol has an associated probability representing its likelihood of occurrence. For example, if the symbols are characters in text, then a common character such as "e" might have an associated probability of 12%.
  - ✓ Special cases, such as the empty set, need to be handled separately.

#### 动机(Motivation)

- 许多作者在撰写论文时非常重视结构安排,但却没有 把这种结构清晰地传达给读者。
- 论文的各个部分不仅应按逻辑顺序排列,还应明确传 达这种逻辑。
- 引言应概述研究内容,并提示论文结构,但这远远不够!

#### 写作中的结构动机(Motivation)

- ▶结构不仅要有,还要让读者"看得见"
  - 很多作者重视结构设计,却没有向读者清楚地展示结构、
  - 论文的各个部分不仅应按逻辑顺序排列,还应明确 传达这种逻辑。

#### 写作中的结构动机(Motivation)

·每一部分**开头和结尾的简要总结**很有帮助,**段与段之** 

间的衔接句也同样重要。

✓ Together these results show that the hypothesis holds for linear coefficients. The difficulties presented by non-linear coefficients are considered in the next section.

• 告诉读者你要说什么 → 说出来 → 告诉他们你说完了

#### 动机(Motivation)

- 常见错误:
  - 定义/定理/算法出现太早,缺乏背景或动机。
  - 作者熟悉的内容, 读者未必知道 —— 不能跳步!
- 建议:
  - 在每个关键步骤提醒读者: 这是什么? 为什么重要? 接下来做什么?
  - 随时思考: "读者此刻知道什么?能理解下面的内容吗?

#### 平衡性 (Balance)

• 平衡 = 对每个话题给予合理的篇幅与深度

#### 無型不平衡例子

- 用一段话带过旧算法,却用七页讲自己的改进。
- 给略述的算法画了几十张图,却没做深入分析。

#### 语态 (Voice)

- 应避免过度使用被动语态
- 主动语态通常更自然、更易读。

- X The following theorem can now be proved.
- We can now prove the following theorem.

### 谨慎使用"学术式"动词

- 常见误区:以为 "utilize / perform" 更科学,实际上,这类词往往可以直接删除或换成简单说法
- X Tree structures can be utilized for dynamic storage of terms.
- Terms can be stored in dynamic tree structures.
- X Local packet transmission was performed to test error rates.
- Error rates were tested by local packet transmission.

#### 其他常见"伪学术动词":

achieved, carried out, conducted, occurred, effected.

#### "We"的用法建议

摘要或引言中,使用"we"有助于区分你自己的贡献与已有的研究成果。

- X It is shown that stable graphs are closed. (谁证明的?)
- We show that stable graphs are closed.
- 💢 It is hypothesized that... (是谁提出的假设?)
- ✓ We hypothesize that...
- ★ This paper shows…" (文章没有意识)

#### "We"的用法建议

➤但在某些情况下使用 "we" 是不合适的:

- X When we conducted the experiment, it showed that our conjecture was correct.
- The experiment showed that our conjecture was correct.

• 实验结论不应因人而异

#### 单元内容

- ▶论文写作为何重要
- ▶好的写作风格
- ▶具体风格
  - 标题与小标题
  - 开头段落
  - 段落
  - 变化
  - 歧义
  - 句子结构
  - 时态
  - 强调
  - 定义
  - 用词

### 标题

#### ▶好标题

- 简洁明了、信息明确
- 避免使用词句冗长

- X A New Signature File Scheme Based on Multiple-Block Descriptor Files for Indexing Very Large Data Bases
- ✓ Signature File Indexes Based on Multiple-Block Descriptor Files
- X An Investigation of the Effectiveness of Extensions to Standard Ranking Techniques for Large Text Collections
- ✓ Extensions to Ranking Techniques for Large Text Collections

### 标题

- ▶好标题
  - 不要过短或内容空泛

- X Huffman Coding for Databases.
- ☑ Limited-Memory Huffman Coding for Databases of Textual and Numeric Data.

### 标题

- ▶好标题
  - 尽量使用具体术语而非笼统表述
  - 标题准确性 > 吸引力

X Strong Modes Can Change the World!

### 小标题

#### ▶小标题应体现论文结构层次

- Lists and Trees
  - <a href="Lists">Lists</a>
  - **Trees**
  - X Other Data Structures

- Index Organizations
  - B-trees
  - X B-trees Indexes

▶第一句话应直接、清晰。

- X Trees, especially binary trees, are often applied—indeed indiscriminately applied—to management of dictionaries.
- Dictionaries are often managed by a data structure such as a tree, but trees are not always the best choice for this application.

- ▶不要用"这篇论文不做什么"开头
- X This paper does not describe a general algorithm for transactions.
- 这样的开头模糊且缺乏吸引力,直到后文读者才发现这篇论文实际描述的是某种特殊情况的算法。
- General-purpose transaction algorithms guarantee freedom from deadlock but can be inefficient. In this paper we describe a new transaction algorithm...
- 要先建立"为什么你要提出新方法"的背景,而不是"否定你没做的"

- ▶避免空泛陈述或"套话"开头
- X Use of digital libraries is increasingly common.
- X It is important that disk access cost be reduced.

- ☑ Digital libraries provide fast access to large numbers of documents.
- Query processing can involve many disk accesses.

- ▶给出研究的背景
- X Underutilization of main memory impairs the performance of operating systems.
- Operating systems are traditionally designed to use the least possible amount of main memory, but such design impairs their performance.

#### 第二种写法更好:

给出了背景(即操作系统通常不怎么用内存)

语气更积极明确,而不是负面模糊的陈述

>要注意区分**已有知识的描述**和**论文自身贡献的描述**。

- X Many user interfaces are confusing and poorly arranged. Interfaces are superior if developed according to rigorous principles.
- Many user interfaces are confusing and poorly arranged.

  We demonstrate that interfaces are superior if developed according to rigorous principles.

### 段落 (Paragraph)

- > 一个段落应专注于一个话题或论点
- 第一句:提出要点(主题句)
- 后续句: 补充或举例说明
- 最后一句: 比中间句更具影响力, 因此要注意句子顺序。

- ☆ 如果一个段落太长,可能意味着你混合了多个论点,应尝试拆分
- ⚠ 警惕:整篇文章段落长度过于一致会显得单调无趣

## 段落 (Paragraph)

>段落之间应逻辑连贯,避免语境丢失

- This algorithm... / his proof...
- ☑ The fast sorting algorithm... / Harvey's proof...

• 通过重复使用关键词或短语来连接段落

### 变化 (Variation)

- ▶在组织结构、句式、段落长度以及用词选择方面保持 多样性,是吸引读者注意力的一种有效手段。
  - X The system of rational numbers is incomplete. This was discovered 2000 years ago by the Greeks. The problem arises in squares with sides of unit length. The length of the diagonals of these squares is irrational. This discovery was a serious blow to the Greek mathematicians.
  - ✓ The Greeks discovered 2000 years ago that the system of rational numbers is incomplete. The problem is that some quantities, such as the length of the diagonal of a square with unit sides, are irrational. This discovery was a serious blow to the Greek mathematicians.

第二个版本: 尽管最后一句没有修改, 它的表达效果却更好。

- > 歧义是写作中的常见问题,因为你知道你想说什么
  - ,反而难以察觉读者可能产生的误解
    - 审查每个句子的主语和代词是否指向明确
    - 用更直接、具体的表达来避免误读

➤确保代词如 "it"、"this" 和 "they" 有明确的指代 对象

- X The compiler did not accept the program because **it** contained errors.
- ✓ The program did not compile because it contained errors.

问题: "it" 是指程序还是编译器? 句子含糊不清。

- ▶另一个与 "it" 有关的问题是: 使用过多。
- X The machine crashed and it was necessary to reboot it.
- The machine crashed and a reboot was necessary.

• 第一句话虽不含歧义,但 "it"被用在了两个不同的意义上。 在不会使句子太笨拙的前提下,应尽量使用更具体的词汇。

- ▶过早使用代词也会引起理解上的困难。
- >> When it was first developed, recursive compilation was impractically slow and required too much memory.
- When recursive compilation was first developed, it was impractically slow and required too much memory.

• 第一句话虽不含歧义,但 "it"被用在了两个不同的意义上。 在不会使句子太笨拙的前提下,应尽量使用更具体的词汇。

### 句子结构(Sentence structure)

>避免嵌套句,即在句中夹杂过多定义或解释

- X In the first stage, the backtracking tokenizer with a two-element retry buffer, errors, including illegal adjacencies as well as unrecognized tokens, are stored on an error stack for collation into a complete report.
- √ The first stage is the backtracking tokenizer with a two-element retry buffer. In this stage possible errors include illegal adjacencies as well as unrecognized tokens; when detected, errors are stored on a stack for collation into a complete report.

原句把主要信息(如何处理错误)和定义混在一起。

### 句子结构(Sentence structure)

>条件表达应连贯,避免支离破碎。

- X If the machine is lightly loaded, then speed is acceptable whenever the data is on local disks.
- ☑ If the machine is lightly loaded and data is on local disks, then speed is acceptable.

#### 句子结构(Sentence structure)

▶注意修饰语的位置错误。

- We collated the responses from the users, which were usually short, into the following table.
- The users' responses, most of which were short, were collated into the following table.

### 重复与平行结构(Repetition and Parallelism)

▶ 互为补充的概念,应使用**平行结构**进行解释,否则读

#### 者难以看出它们之间的对应关系

- X In SIMD, the same instructions are applied simultaneously to multiple data sets, whereas in MIMD different data sets are processed with different instructions.
- ✓ In SIMD, multiple data sets are processed simultaneously by the same instructions, whereas in MIMD multiple data sets are processed simultaneously by different instructions.

### 重复与平行结构(Repetition and Parallelism)

- 缺乏平行结构会导致歧义
- X The performance gains are the result of tuning the low-level code used for data access and improved interface design.
- The performance gains are the result of tuning the low-level code used for data access and **of** improved interface design.
- 将较长的部分放到句子后面更清晰
- ✓ The performance gains are the result of improved interface design and of tuning the low-level code used for data access.

- ▶在科学写作中,大多数文本使用的是**过去时**或**现在时**
- 现在时
  - 用于表达永恒真理:
     the algorithm has complexity O(n)
  - 描述文章本身的内容:
    - x related issues are discussed below
    - related issues will be discussed below

- 过去时用于描述已完成的工作和研究结果:
  - X the ideas are tested by experiment
  - the ideas were tested by experiment

- > 有时需要混合使用现在时和过去时
- 理论为现在时,实验观察为过去时:
  - Although theory suggests that the Klein algorithm has asymptotic complexity O(n2), in our experiments the trend observed was O(n).

- ▶有时需要混合使用现在时和过去时
- 理论为现在时,实验观察为过去时:
  - Although theory suggests that the Klein algorithm has asymptotic complexity O(n2), in our experiments the trend observed was O(n).

- ▶在引用文献时,现在通常更合适,但在某些上下文中也可能需要使用过去时。
  - ✓ Willert (1999) shows that the space is open .
  - ✓ Haast (1986) **postulated** that the space is bounded, but Willert (1999) has since shown that it is open.
- 除非在撰写结论部分, 科学写作中很少使用将来时。

### 强调(Emphasis)

- ▶句子的结构本身会**隐含**地突出某些词语。改变句子的结构,也会改变语义的重点。
- X A static model is appropriate because each item is written once and read often.
- A static model is appropriate because each item is only written once but is **read often**.

### 强调(Emphasis)

▶首次引入关键词时可以斜体

The data structure has two components, a *vocabulary* containing all of the distinct words and, for each word, a hit list of references.

### 定义(Definitions)

▶术语、变量、缩写和首字母缩略词,在首次使用时应 当加以定义或解释

- X We use homogeneous sets to represent these events.
- We use *homogeneous* sets to represent these events.
- To represent these events we use homogeneous sets, whose members are all of the same type.

### 定义(Definitions)

- >陌生概念, 多种方式解释会更清楚
- Compaction, in contrast to compression, does not preserve information; that is, compacted data cannot be exactly restored to the original form.
- → 使用对比 + 举例的方法, 有助于澄清概念。

▶应优先使用简洁、直接的词汇,而非冗长、绕圈子的表达方式,这样写出来的文字会更有力、更明确。

- 用 "begin" 替代 "initiate"
- 用 "first"、"second" 替代 "firstly"、"secondly"
- 用 "part" 替代 "component"
- 用 "use" 替代 "utilize"

- ▶避免抽象、含糊或太宽泛的词。
- X The analysis derives information about programs.
- → "information"可能指函数分析、bug报告、复杂度等,含义过泛
- The analysis estimates the resource costs of programs.

• 避免过度使用常见但模糊的术语,如: important, intelligent, method, paradigm, performance, semantic...

- ▶ "困难"应该更具体
  - 如果某事 "difficult to compute" (难以计算),
     那是因为它慢?占内存多?需要高精度?还是其他原因?
    - →应用更精确的描述。
- ➤efficient 也是含糊词,应尽量具体说明效率在哪方面

- ▶避免为避免重复而滥用同义词
- X The database executes on a remote machine to provide better security for **the system** and insulation from network difficulties.
- ✓ The database executes on a remote machine to provide better security for **the database** and insulation from network difficulties.
- 科技写作中, "不要重复用词"这条规则并不适用。术语应清晰, 而非刻意变换。

# 限定词(Qualifiers)

- > 不要在一句话中堆叠多个限定词
- 限定词(如 might, may, perhaps, possibly, likely, likelihood, could) 最多使用一个, 避免文本软弱
- X It is perhaps possible that the algorithm might fail on unusual input.
- ☑ The algorithm might fail on unusual input.
- It is possible that the algorithm would fail on unusual input.

### 限定词(Qualifiers)

- ▶避免使用very、quite等无意义修饰语
- 这些词在技术写作中基本没有实质意义。例如: "very fast" 的算法,是否就比"fast"的算法更好?

- X There is very little advantage to the networked approach.
- $ule{\hspace{-0.1cm}\hspace{-0.$

- ➤ Which, That
- · 许多作者在本应使用 "that" 的地方使用了 "which"。
  - →除非无法用 "that" 替代,应优先使用 "that"。
- X There is one method which is acceptable.
- There is one method that is acceptable.
- There are three options, of which only one is tractable.

- ➤ Alternate / Alternative / Choice
  - alternate: 交替的、另一的
  - · alternative: 可选择的方案
  - · choice: 有多个选项中的选择
  - → 如果只有一个 "alternative" , 其实并没有 "choice" 。

- ➤ Basic / Fundamental
  - · basic: 基础的、初级的
  - fundamental: 根本的、本质的
  - → 如果一个结果被称为 "basic" , 那它应当确实是 "初级的"

- Fast / Quickly / Presently / Timely / Currently
  - fast: 快速
  - quickly: 迅速, 但不一定 "马上"
  - presently: 很快、即将
  - currently: 当前
  - timely: 时机恰当的(与速度无关)

Usual	Other	说明
alternative	alternate	alternative 表示"可选择项";alternate 表示"交替的"。
comparable	comparative	comparable 表示"可比较的";comparative 通常指"比较的"或语法中"比较级"。
complement	compliment	complement 表示"补足、补语";compliment 是"赞美"。
dependent	dependant	dependent 是形容词/美式拼写("依赖的"); dependant 是英式拼写中的名词(依赖的人)。
descendant	descendent	descendant 是正确的用法,意指"后代";descendent 较罕见或视为错误。
discrete	discreet	discrete 表示"离散的、不连续的"; discreet 表示"谨慎的"。
emit	omit	emit 是"发出(光、热)";omit 是"省略、忽略"。
ensure	insure	ensure 是"确保";insure 通常用于"投保"。
ensure	assure	ensure 是"确保事件发生";assure 是"向人保证"。
excerpt	exert	excerpt 是"摘录";exert 是"施加(力量/影响)"。

Usual	Other	说明
foregoing	forgoing	foregoing 表示"上述的"; forgoing 是 forgo(放弃)的现在分词。
further	farther	further 多用于抽象延伸;farther 用于物理距离。
elusive	illusive	elusive 表示"难以捉摸的"; illusive 表示"虚幻的、不真实的"。
manyfold	manifold	manyfold 较旧或不常用;manifold 表示"多方面的"。
omit	emit	同左侧解释。此处是反向混淆。
partly	partially	partly 表示"部分地"(强调数量或范围); partially 侧重"不完全"。
principle	principal	principle 是"原则";principal 是"校长、主要的"。
simple	simplistic	simple 是"简单的";simplistic 含贬义,表示"过于简化"。
solvable	soluble	solvable 是"可解决的";soluble 是"可溶解的"。
stationary	stationery	stationary 是"静止的";stationery 是"文具"。

#### 用词重复(Overuse of Words)

▶当一个词或短语重复出现(或出现倒装形式),读者会觉得冗余,降低阅读体验。

- X Ada was used for this project because the underlying operating system is implemented in Ada.
- Ada was used for this project because it is the language used for implementation of the underlying operating system.

• 但重复有时是必要的,尤其是为了消除歧义。

#### 用词重复(Overuse of Words)

▶避免一个词在不同语境中重复

- X Values are stored in a set of accumulators, each initially set to zero.
- ✓ Values are stored in a set of accumulators, each initialized to zero.

#### 用词重复(Overuse of Words)

- > 一些用词重复的例子
- first of all ->first
- for the purpose of-> for
- Semantic meaning ->meaning
- completely random-> random
- a number of ->several
- such as . . . etc.-> such as ...

谢谢!!