Chinese Systematic Naming Saturated Chain Hydrocarbon (SCH) by Graph Algorithms

George

@bit.edu.cn MonkeyHouse.Info

#### Outline

Overview
Transformation
Algorithms and its Optimizing
Conclusions and Future Work

# 1. OverviewProblems came from Learning

#### 1. Overview

- Problems came from Learning
- Naming Manually
  - Tedious
  - Low Efficient

#### 1. Overview

- Problems came from Learning
- Naming Manually
  - Tedious
  - Low Efficient
- Abstracting of Organic Compounds
  - SCH becomes Tree Network
  - Chemistry Problem goes into Graph

#### Rules of Chinese Systematic Naming 1. Longest Main Chain (MC) 2. "Nearest Smallest" 3. Simple before Complex $CH_3$ C

CH<sub>3</sub>

 $CH_3 - C - CH_3$ 

CH<sub>3</sub>

CH<sub>3</sub>

#### Rules of Chinese Systematic Naming 1. Longest Main Chain (MC) $CH_3$ 2. "Nearest Smallest" 3. Simple before Complex CH<sub>3</sub> C $-CH_3$ $CH_3$ $CH_3$ ĈH<sub>3</sub> $CH_3$ С CH<sub>3</sub> CH<sub>3</sub> CH<sub>3</sub> C \_CH<sub>3</sub>

#### Rules of Chinese Systematic Naming 1. Longest Main Chain (MC) CH<sub>3</sub> 2. "Nearest Smallest" 3. Simple before Complex С CH<sub>3</sub> $CH_3$ CH<sub>3</sub> $CH_3$ CH<sub>3</sub> С $CH_3$ CH<sub>3</sub> CH<sub>3</sub> CH<sub>3</sub> C \_CH<sub>3</sub>

#### Rules of Chinese Systematic Naming 1. Longest Main Chain (MC) CH<sub>3</sub> 2. "Nearest Smallest" 3. Simple before Complex С \_CH<sub>3</sub> $CH_3$ CH<sub>3</sub> CH<sub>3</sub> $CH_3$ С $CH_3$ CH<sub>3</sub> CH<sub>3</sub> $CH_3 - C - CH_3$ CH2

2,2,4,4-四甲基-3,3-二(1,1-二甲基乙基)戊烷



# 2. TransformationOrganic Compounds→Graph



# Organic Compounds→Graph SHC→Tree Network



# 2. Transformation Organic Compounds→G SHC→Tree Network

*Chemistry* Organic Compounds

*Mathematics* Graph Theory

## Longest MC→Diameter of Graph Best Naming→Simulation & Comparing

Longest MC→Diameter of Graph
Best Naming→Simulation & Comparing
Grouping Side Chains→Isomorphic of Graph



- Longest MC→Diameter of Graph
- Best Naming→Simulation & Comparing
  - Grouping Side Chains→Isomorphic of Graph
- SHC $\rightarrow$ Tree Network (TN)
- Longest MC→Diameter of TN
- Grouping Side Chains (SC)→IsomorphiTrees

Longest MC→Diameter of Gra
Best Naming→Simulation & C



Grouping Side Chains—Isomorphic of Caph

*Mathematics* Graph Problem

- SHC→Tree Network (TN)
  Longest MC→Diameter of TN
- Grouping Side Chains (SC)→Is





Best Naming Simu. & Comp. (P Class)

Longest MC Diameter of TN (P Class)

Best Naming Simu. & Comp. (P Class)

Longest MC Diameter of TN (P Class) Grouping SC Isomorphic of Trees (P Class)













Best Naming Simu. & Comp. (P\_Class)

Longest MC Diameter of TN (P Class) Grouping SC Isomorphic of Trees (P Class)



3,3,5,5-四甲基-4,4-二(1,1-二甲基乙基)辛烷

#### $O(m^2) + O(n^2) + O(n^3) + O(m^3) = O(n^3)$ 3.1. Algorithms – Time Complexity

Procedure Naming(G) if 过程Naming(G)是第一次调用 then 对G中所有取代基Z进行复杂性比较的预处理 ・Sorting SCs の(m <sup>2</sup> ) if G是烃 then			
if 过程Naming(G)是第一次调用 then 对G中所有取代基Z进行复杂性比较的预处理 if G是烃 then			
then 对G中所有取代基Z进行复杂性比较的预处理 if G是烃 then			
对G中所有取代基Z进行复杂性比较的预处理 if G是烃 then			
if G是经 then			
then			
(-2) + (2)			
$\mathbb{D} \leftarrow FindD(\mathbb{G})$ $\blacktriangleright$ •Get Dia. of TN $O(n^2) + (2m-1) \times O(n) = O(n^2)$	$O(n^2)^+(2m-1) \times O(n) = O(n^2)$		
else			
$D \leftarrow FindDJ(G)$ J $\rightarrow$ •Get the MC with most SCs $2m \times n \times O(n)$			
将 $Count(S)$ 最大的直径 $S \in D$ 存入集合 $\mathbb{R}$			
for each $S \in \mathbb{R}$ do $= O(n^3)$			
for each $\mathbb{Z} \in \mathbb{P}_S$ do	1		
if $Hash(Rank(\mathbb{Z})) = NULL$ then			
<u>将Z的地址存入Hash(Rank(Z))</u> ・Naming SC Recursively			
for each $\mathbb{Z} \in \mathbb{K}$ do			
对侧链Z调用Naming(Z)进行系统命名AZ			
for each $\mathbb{S} \in \mathbb{R}$ do			
for each ℤ ∈ K do •Connecting Numbers and SC	Cs		
将Hash(Rank(ℤ))∩S的标号与Az连接后,存入该直径的侧链系统命名集合Ms Sorting SCs O(3)			
通过 $Compare(\mathbb{Z}_1, \mathbb{Z}_2)$ 对侧链 $\mathbb{Z} \in \mathbb{K}$ 的复杂性进行堆排序			
将命名开始编号一端距离最简单侧链最近的直径S∈R存入集合R/ ・Choosing the Best Namin	ng i		
$\mathbb{S} \in \mathbb{R}'$ 中,侧链序号和最小的命名为 $\mathbb{S}_g$			
将Ms,中对侧链的命名按照顺序输出并用短线连接,最后输出碳原子数为card(D)的直链。Output in Natural angua	ade		
烷烃的习惯命名	ge		

#### $O(m^2) + O(n^2) + O(n^3) + O(m^3) = O(n^3)$ 3.1. Algorithms – Time Complexit

算法 3 饱和链烃和一价饱和链烃基系统命名算法			41	
▷Procedure Naming(G)			appx. 1k	
if 过程Naming(G)是第一次调用				
then	•Sorting SCs	$O(m^2)$	carbons/sec	
对G中所有取代基Z进行复杂性比较的预处理	3			
if G是烃		Contraction of the second		
then		O(2) + O	1)	
$\mathbb{D} \leftarrow FindD(\mathbb{G})$ • Get Dia	. of TN	$O(n^{2})+(2m)$	$(-1) \land O(n) = O(n^2)$	
else				
$\mathbb{D} \leftarrow FindDJ(\mathbb{G})$	<ul> <li>Get the MC wit</li> </ul>	h most SCs	$2m \times n \times O(n)$	
将 $Count(S)$ 最大的直径 $S \in D$ 存入集合 $\mathbb{R}$		1 1 2 3 2 - 1		
for each $\mathbb{S} \in \mathbb{R}$ do	1		$=O(n^{3})$	
for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do		and the second second		
if $Hash(Rank(\mathbb{Z})) = \text{NULL then}$	Crouning SC		and the second second	
$\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$	L .arouping sc		in the second	
将Z的地址存入Hash(Rank(Z))	Naming SC R	Recursively		
for each $\mathbb{Z} \in \mathbb{K}$ do				
对侧链Z调用Naming(Z)进行系统命名Az				
for each $\mathbb{S} \in \mathbb{R}$ do	- 1			
for each $\mathbb{Z} \in \mathbb{K}$ do		•Connecting	Numbers and SCs	
将Hash(Rank(ℤ))∩S的标号与Az连接后,存入	该直径的侧链系统命名集合Ms	Sorting S	$C_{S} \cap (\pi\pi^3)$	
通过 $Compare(\mathbb{Z}_1, \mathbb{Z}_2)$ 对例键 $\mathbb{Z} \in \mathbb{K}$ 的复杂性进行堆	排序		$O(m^2)$	
将命名开始编号一端距离最简单侧链最近的直径S	∈ <b>ℝ存入集合</b> ℝ′	•Choosing the Best Naming		
$S \in \mathbb{R}'$ 中,侧链序号和最小的命名为 $S_g$		and the state		
将Ms,中对侧链的命名按照顺序输出并用短线连接	長,最后输出碳原子数为card(D)的直链	•Output in	Natural Language	
烷烃的习惯命名			jj.	

## $O(m^2) + O(n^2) + O(n^3) + O(m^3)$ 3.2. Optimized Algorithms = $O(n^3)$

算法 3 饱和链烃和一价饱和链烃基系统命名算法	Saller and a set of the second
>Procedure Naming(G) if <b>过程</b> Naming(G) <b>是第一次调用</b>	
then •Sorting SCs	$O(m^2)$
对G中所有取代基型进行复杂性比较的预处理	
if G是烃	
then Cat Dia of The	$O(-2) + (2 - 1) \times O(-2) - O(-2)$
$\mathbb{D} \leftarrow FindD(\mathbb{G})$ • Get Dia. of IN	$O(n^2)^+(2m-1) \wedge O(n) = O(n^2)$
else	
$\mathbb{D} \leftarrow FindDJ(\mathbb{G})  \mathbb{J} \qquad \mathbb{P}  \mathbb{P}  \mathbb{P}$	with most SCs $2m \wedge n \wedge O(n)$
将 $Count(S)$ 最大的直径 $S \in D$ 存入集合 $\mathbb{R}$	-O(-3)
for each $\mathbb{S} \in \mathbb{R}$ do	$=O(n^{2})$
for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do	The second se
if $Hash(Rank(\mathbb{Z})) = NULL$ then	SC
$\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$	30
將ℤ的地址存入Hash(Rank(ℤ)) ・Naming S	C Recursively
for each $\mathbb{Z} \in \mathbb{K}$ do	
对侧链Z调用Naming(Z)进行系统命名AZ	
for each $\mathbb{S} \in \mathbb{R}$ do	
for each $\mathbb{Z} \in \mathbb{K}$ do	•Connecting Numbers and SCs
将 $Hash(Rank(\mathbb{Z}))$ ∩S的标号与 $A_Z$ 连接后,存入该直径的侧链系统命名集合 $M_S$	Sorting SCs (
通过 $Compare(\mathbb{Z}_1, \mathbb{Z}_2)$ 对侧链 $\mathbb{Z} \in \mathbb{K}$ 的复杂性进行堆排序	$O(m^3)$
将命名开始编号一端距离最简单侧链最近的直径S∈ ℝ存入集合ℝ′	•Choosing the Best Naming
$\mathbb{S} \in \mathbb{R}'$ 中,侧链序号和最小的命名为 $\mathbb{S}_g$	the second states of the secon
将M <sub>Sg</sub> 中对侧链的命名按照顺序输出并用短线连接,最后输出碳原子数为card(D)	nit - Output in Natural Language
院经的习惯命名	output in Natural Edilguag

## $O(m^2) + O(n^2) + O(n^3) + O(m^3)$ 3.2. Optimized Algorithms = $O(n^3)$



#### $O(m^2) + O(n^2) + O(n^2) + O(m^2 \log m)$ 3.2. Optimized Algorithms = $O(n^2 \log m)$



#### $O(m^2) + O(n^2) + O(n^2) + O(m^2 \log m)$ 3.2. Optimized Algorithms $n^2\log n$ 算法 3 饱和链烃和一价饱和链烃基系统命名算法 Procedure Naming(G) appx. 10k if 过程Naming(G)是第一次调用 •Sorting SCs then carbons/sec 对G中所有取代基Z进行复杂性比较的预处理 if G是烃 then Get MC with longest dia. FindD&Count(G and most SCs else FindD&Count(G on the same time for each $S \in \mathbb{R}$ do for each $\mathbb{Z} \in \mathbb{P}_{\mathbb{S}}$ do if $Hash(Rank(\mathbb{Z})) = NULL$ then Grouping SC $\mathbb{K} = \mathbb{K} \cup \{\mathbb{Z}\}$ Naming SC Recursively 将Z的地址存入Hash(Rank(Z)) for each $\mathbb{Z} \in \mathbb{K}$ do 对侧链Z调用Naming(Z)进行系统命名Az for each $S \in \mathbb{R}$ do •Connecting Numbers and SCs for each $\mathbb{Z} \in \mathbb{K}$ do 將 $Hash(Rank(\mathbb{Z}))$ ∩S的标号与 $A_Z$ 连接后,存入该直径的侧链系统命名集合 $M_s$ •Sorting SCs $O(m^2 logm)$ •Choosing the Best Naming 通过 $Compare(\mathbb{Z}_1, \mathbb{Z}_2)$ 对侧链 $\mathbb{Z} \in \mathbb{K}$ 的复杂性进行堆排序 将命名开始编号一端距离最简单侧链最近的直径S∈ ℝ存入集合ℝ′ $S \in \mathbb{R}'$ 中, 侧链序号和最小的命名为 $S_a$ 將Ms,中对側链的命名按照順序输出并用短线连接,最后输出碳原子数为card(型)的直链 ● Output in Natural Language 院经的习惯命名

4.1. Conclusions SHC→Tree Network in Graph Theory Getting MC and Grouping SCs→Diameters and isomorphism of TN Give Chinese Systematic Naming Algorithm (CSNA) based on Diameterand Isomorphism Algorithms **Optimize the Algorithm to Get Better Performance CSNA: Efficient and Fast** 

#### 4.2. Future Work

Accurate Computing Complexity Estimation
Pattern Matching :

Generating TNs from Expressions

Generalization

# Thanks

## Thanks for Your Attention