

# Một số vấn đề và phương pháp học tập và nghiên cứu khoa học

Hồ Tú Bảo

Japan Advanced Institute  
of Science and Technology

Vietnamese Academy  
of Science and Technology

## Phần 1

### VỀ VIỆC HỌC CAO HỌC VÀ NGHIÊN CỨU KHOA HỌC

(from talks at HCMUT 2007, ICT-Hanoi 2007, and some writings)

2

## Bản chất của đào tạo sau đại học

### ■ Bản chất của đào tạo thạc sĩ là **học**

- Học là việc chuyển tri thức con người đã biết thành tri thức của các cá nhân hoặc tổ chức.
- Đại học: học các tri thức chung của nghề; Thạc sĩ: học các tri thức chuyên sâu của nghề.
- Thạc sĩ là người tinh thông nghề nghiệp (master, étude approfondie).

### ■ Bản chất của đào tạo tiến sĩ là **nghiên cứu**

- Nghiên cứu là việc tìm và tạo ra các tri thức mới và có ý nghĩa bởi các cá nhân hoặc tổ chức.
- Tiến sĩ là người biết làm nghiên cứu, và chủ yếu làm việc nghiên cứu.



3

## Bản chất của đào tạo thạc sĩ là học

### ■ Chương trình thạc sĩ phổ biến trên thế giới

- Học hai năm với tín chỉ
- Năm đầu chủ yếu học các môn cần thiết (khoảng 10 môn, phần lớn tự chọn)
- Năm thứ hai chủ yếu cho việc rèn luyện
  - seminar, reading, hoạt động của lab
  - làm đề tài nghiên cứu, viết và bảo vệ luận văn.



### ■ Chương trình thạc sĩ phổ biến của ta

- Phần lớn thời gian cho các môn học trên lớp (khoảng 20 môn)
- Chưa dùng hệ tín chỉ
- Ít thời gian cho rèn luyện và làm luận văn
- Ít rèn khả năng tự học
- Tiêu chí và cách đánh giá chưa thích hợp (luôn yêu cầu cái mới)?



4

## Bản chất của đào tạo tiến sĩ là nghiên cứu

- **Đòi hỏi cơ bản**
  - Biết xác định được vấn đề nghiên cứu có ý nghĩa
  - Biết giải quyết vấn đề
  - Biết viết bài và trình bày.
- **Rất nhiều luận án làm ứng dụng thay vì nghiên cứu cơ bản hay ứng dụng**
- **Kết quả chủ yếu công bố ở các tạp chí hay hội nghị trong nước.**
- **Xa điều kiện của Bộ GD-ĐT:**
  - 01 bài báo tạp chí quốc tế, một vài bài hội nghị quốc tế
  - 01 bài báo tạp chí trong nước
- **Thách thức lớn! 1995-2004:**
  - 5259 GS & PGS
  - 3236 (800) bài tạp chí quốc tế



5

(<http://vietnamnet.vn/khoahoc/vande/2006/01/532815>)

## Some Ph.D. thesis

(Tạp chí Tia Sáng, 18.10.2007)

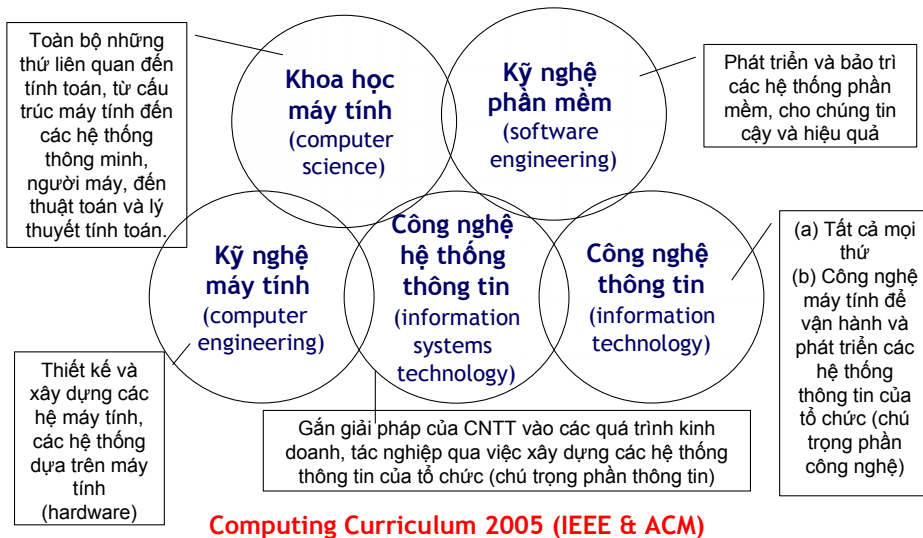
- “Nhận thức của công chức hành chính về việc sắp xếp lại bộ máy của cơ quan chuyên môn trực thuộc Ủy ban nhân dân cấp tỉnh, thành phố”
- “Nhận thức của thanh niên nông thôn về chất lượng cuộc sống gia đình hiện nay”
- “Nghiên cứu nhu cầu điện ảnh của sinh viên”
- “Phát huy vai trò của tri thức ngành y tế Việt Nam trong công cuộc đổi mới”.
- “Lịch sử phát triển giáo dục-đào tạo ở An giang (1975 - 2000)”



<http://www.hed.edu.vn/TrangChu/LuanAnTienSi/TomTatNhungDiemMoi/>

6

## Chương trình và sách giáo khoa ICT

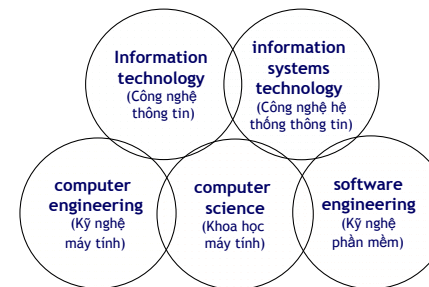


CC 2005: [http://www.computer.org/portal/cms\\_docs\\_ieeeecs/ieeeecs/education/cc2001/CC2005-March06Final.pdf](http://www.computer.org/portal/cms_docs_ieeeecs/ieeeecs/education/cc2001/CC2005-March06Final.pdf)  
Before 1990s, in North America: computer science, electrical engineering, and information systems

7

## Chương trình và sách giáo khoa ICT

“Sinh viên không chỉ cần học những thứ hiện nay đã được biết, mà cần cả cách làm sao giữ cho kiến thức của mình được cập nhật. Các công cụ dựa trên công nghệ để thu thập tri thức cần phải trở thành những yếu tố trung tâm của giáo dục cho họ, và chương trình cần được xây dựng để sinh viên **học cách học**” \*



- Khả năng tự học là cốt yếu trong ICT
- Phụ thuộc chương trình, thầy và trò.



\* “Peril and Promise: Higher Education in Developing Countries”, World Bank & UNESCO

8

## Khoa học và Công nghệ

- **Khoa học** là việc khảo sát các hiện tượng tự nhiên và xã hội để tìm tri thức mới.
- **Công nghệ** là cách dùng các tri thức khoa học và vật liệu để đạt mục tiêu làm sản phẩm (“technology is not about tools, it deals with how humans work”, Peter Drucker).
- Thay đổi khắp nơi ở Việt Nam trong các năm 1990s: khoa học → khoa học & công nghệ (Viện KHAVN → Viện KH&CNVN, Bộ Khoa học và Công nghệ, etc.)
- Khoa học và công nghệ rất liên quan đến nhau nhưng là hai thứ khác nhau. KH-CN đang được dùng lẫn vào nhau như một đơn vị của nhận thức (không luôn luôn tốt).
- Việt Nam cần tỷ lệ khoa học IC và công nghệ IC bao nhiêu?



9

## Nghiên cứu cơ bản, nghiên cứu ứng dụng và ứng dụng?

- **Nghiên cứu cơ bản**: Tìm tri thức mới cho các nghiên cứu cơ bản khác hay nghiên cứu ứng dụng
  - Gene finding
  - Mô hình ngôn ngữ tiếng Việt
  - Kernel methods
- **Nghiên cứu ứng dụng**: Tìm tri thức khoa học để giải quyết các vấn đề thực tế
  - Dịch máy Anh-Việt
- **Ứng dụng**: Dùng tri thức đã biết để giải quyết các vấn đề thực tế.
- Trong ICT
  - Nghiên cứu cơ bản có thể nhanh chóng chuyển vào nghiên cứu ứng dụng
  - Nghiên cứu ứng dụng có thể nhanh chóng chuyển thành sản phẩm
  - Ứng dụng có khắp nơi

(<http://vietnamnet.vn/khoahoc/vande/2006/01/532815>)

10

## Nghiên cứu cơ bản, nghiên cứu ứng dụng và ứng dụng?

- “Trong khi không phải mọi đất nước đều cần tiến hành nghiên cứu cơ bản ở nhiều lĩnh vực khác nhau, mỗi đất nước cần phải xem xét các loại nghiên cứu khoa học và công nghệ có thể trực tiếp đóng góp vào sự phát triển của mình.
- ... Có lẽ câu hỏi cần hỏi nhất là: **đâu là mức tối thiểu các hoạt động khoa học và công nghệ cần phải có để đạt được các mục tiêu của quốc gia?**”
- Nghiên cứu cơ bản bao nhiêu phần trăm? Vào vấn đề gì? Lĩnh vực nào? → nên tập trung cho các nghiên cứu làm nền tảng cho nghiên cứu ứng dụng.
- Cần khuyến khích và tổ chức nghiên cứu công nghệ
- Đề cao và ưu tiên cho số đông làm nghiên cứu ứng dụng.

“Peril and Promise: Higher Education in Developing Countries”, World Bank and UNESCO 11

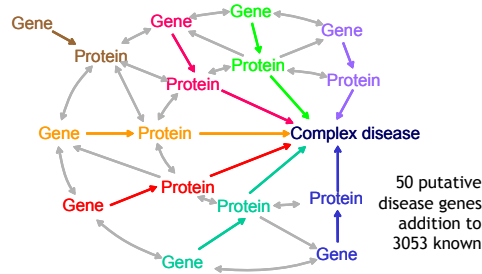
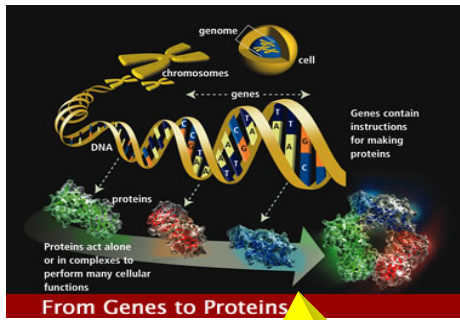
## Nghiên cứu cho nhu cầu ICT của Vietnam

- Các lĩnh vực thiết yếu của ICT: như kỹ thuật mạng, công nghệ phần mềm, an toàn thông tin, trí tuệ nhân tạo, v.v. → công bố quốc tế
- Các lĩnh vực mới, thích hợp và triển vọng
  - Tin sinh học, công nghệ Web, các loại dữ liệu phức tạp ... thay vì các chủ đề đã quá quen thuộc như tập mờ, tập thô, cơ sở dữ liệu quan hệ, ...
- Các lĩnh vực cần cho nhu cầu ở Việt Nam và người Việt phải làm, như:
  - Hành chính điện tử, hạ tầng cơ sở ICT ... vs. thực tại ảo
  - Xử lý văn bản và tiếng nói tiếng Việt
  - Phát hiện đạo văn và cơ sở dữ liệu luận văn



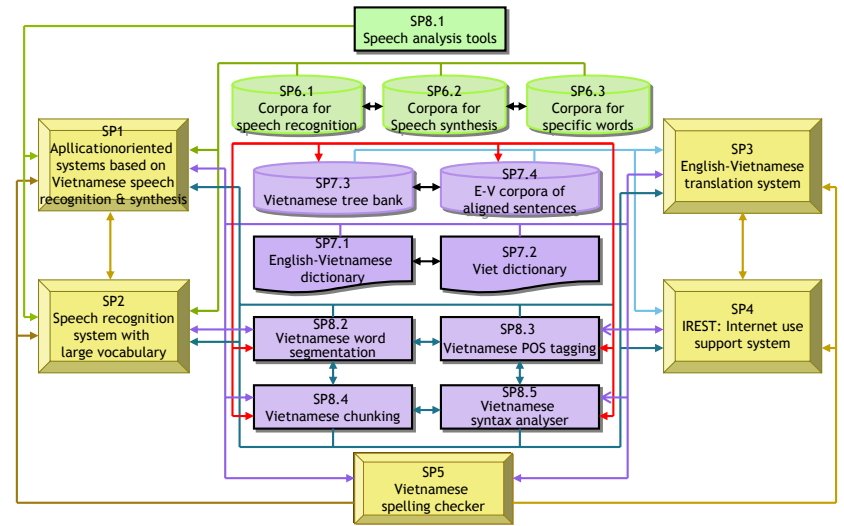
12

# Dự đoán gene gây bệnh và tin y-sinh học



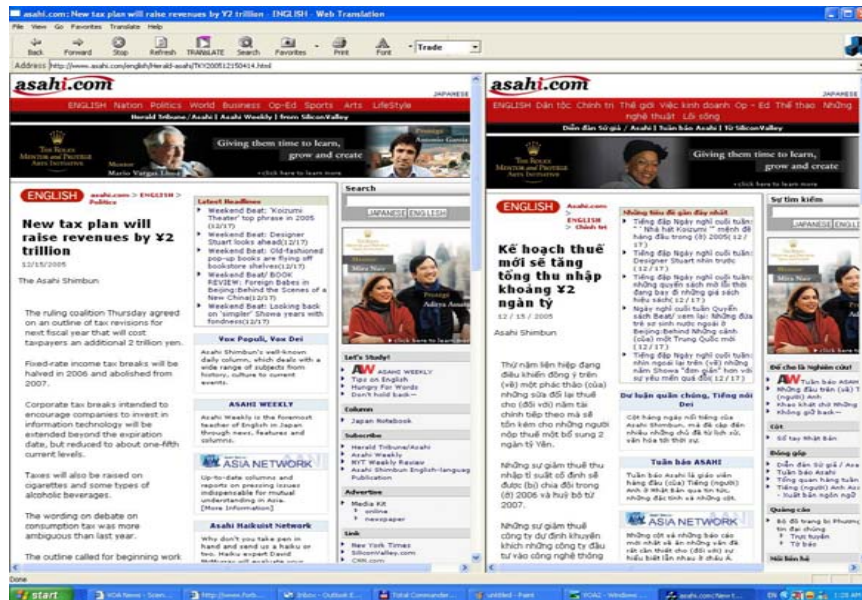
Metabolomics 1400 Chemicals  
 Proteomics 100,000 Proteins  
 Genomics 25,000 Genes

# Hạ tầng cơ sở cho xử lý tiếng nói và văn bản tiếng Việt



National project KC01-01/06-10 on "Vietnamese Language and Speech Processing"

# Dịch máy Anh-Việt



# Hướng đến các công bố quốc tế

- Nghiên cứu cần hướng đến công bố trên các tạp chí và hội nghị quốc tế \*
- Cần khuyến khích và đề cao các nghiên cứu chất lượng cao, và phân biệt giá trị khác nhau của kết quả nghiên cứu (rất cạnh tranh)
- Cần dạy và học phương pháp nghiên cứu khoa học
- Từng bước đạt mục tiêu trên



\* "Văn hóa ngành" trong tiêu chí đánh giá, <http://www.tiasang.com.vn/news?id=1771>

## Hai loại ấn phẩm khoa học chính

- Tạp chí
  - Quốc tế: Google “[computer science journal ranking](#)”
  - Trong nước:
- Hội nghị quốc tế
  - Conferences (hội nghị), symposiums, workshop, forums (hội thảo), congress (đại hội)
  - Quốc tế: Google “[computer science conference ranking](#)”
  - Trong nước: ICT, FAIR, etc.
- Measures: Impact factor, citation
- Trong ICT, hội nghị tốt được đánh giá rất cao (khác với các ngành toán, lý, hóa, sinh học, etc.)

17

## Computer science journal ranking

(Google “[computer science journal ranking](#)”, trích phần AI journals)

- **Premium:** Artificial Intelligence, Artificial Intelligence Review, Computational Linguistics, [IEEE Trans on PAMI, Robotics and Automation, Image Processing, Journal of AI Research, Neural Computation, \[Machine Learning\]\(#\), Intl Jnl of Computer Vision, etc.](#)
- **Leading:** ACM Trans. on Asian Language Information Processing, AI Magazine, Annals of Mathematics and AI, Applied Artificial Intelligence, [Applied Intelligence](#), Artificial Intelligence in Medicine, [IEEE Trans on Neural Networks, Speech and Audio Proc, Systems, Man, & Cybernetics, Part A & B, Intl Jnl on Artificial Intelligence Tools, \[Machine Translation\]\(#\), Neural Networks, Pattern Recognition, etc.](#)
- **Reputable:** Computer Processing of Chinese & Oriental Languages, Intl Jnl of Pattern Recognition & AI, [Computers and Artificial Intelligence, IEEE Transactions on Fuzzy Systems, Journal of Intelligent and Fuzzy Systems, Knowledge Acquisition Jnl, Knowledge-Based Systems, Pattern Recognition Letters, Jnl. of Japanese Soc. of AI, Intelligent Data Analysis, etc.](#)
- **Others:** Canadian Artificial Intelligence, Journal of Advanced Robotics, Journal of Artificial Intelligence in Education, Journal of Artificial Intelligence in Engineering, Automation, and Manufacturing, Journal of Computational Acoustics, Journal of Computational Neuroscience, Journal of Computational Vision, etc.
- **Easy:** WASET (World Academy of Science, Engineering and Technology) , WSEAS

18

## Computer science conference ranking

(Google “[computer science conference ranking](#)”, trích phần hội nghị AI)

- **Rank 1:** IJCAI: Intl Joint Conf on AI, AAAI: American Association for AI National Conference, [ICML: Intl Conf on Machine Learning](#), UAI: Conference on Uncertainty in AI, UM: Intl Conf on User Modeling, [NIPS: Neural Information Processing Systems](#), AGENTS: International Conference on Autonomous Agents [[SIGKDD: ACM Knowledge Discovery and Data Mining](#), [ICDM: IEEE International Conference on Data Mining](#)], etc.
- **Rank 2:** ECAI: European Conf on AI, ECML: European Conf on Machine Learning, GECCO: Genetic and Evolutionary Computation Conference, GP: Genetic Programming Conference, IAAI: Innovative Applications of AI, ICIP: Intl Conf on Image Processing, ICPR: Intl Conf on Pattern Recognition, ICTAI: IEEE conference on Tools with AI, etc. [[COLING: Intl Conf on Computational Linguistics](#), [PAKDD: Pacific-Asia Conf on Know. Discovery & Data Mining](#), [PKDD: European Conf Knowledge Discovery in Databases](#)], etc.
- **Rank 3:** [PRICAI: Pacific Rim Intl Conf on AI](#), [AusAI: Australian Joint Conf on AI](#), etc.
- **Unranked Conferences:** AAMAS: Intl Joint Conf on Autonomous Agents and Multiagent Systems, NFOVIS: IEEE Symp. on Information Visualization, VIS: IEEE Visualization, etc.
- **Not Encouraged (due to dubious referee process):** Intl Multiconferences in Computer Science -- 14 joint int'l confs., SCI: World Multi confs on systemics, sybernetics and informatics , SSGRR: International conf on Advances in Infrastructure for e-B, e-Edu and e-Science and e-Medicine , IASTED conferences, CCCT: International Conference on Computer, Communication and Control Technologies.

19

## Số bài báo tạp chí quốc tế 10 năm

(“made in Vietnam” during 1995-2004)

Areas	# papers	Res. Institutions	Universities	Citation avg
Mathematics	300	144	121	1.4
Theoretical Physics	131	100	31	2.4
Experimental Physics	40	16	24	1.6
Technology	42	25	9	0.8
<b>ICT</b>	<b>38</b>	<b>19</b>	<b>11</b>	<b>1.3</b>
Materials Science	36	9	27	1.8
Medicine	36	28	8	2.3
Chemistry	32	28	4	1.2
Agriculture	23	15	8	1.7
Polymer	19	14	5	1.1
Mechanics	17	15	1	2.6
Social science	14	12	2	0.2
Environmental science	13	13	0	4.1
Biology	10	9	1	1.3
Earth science	9	8	1	0.5
Pharmacy and drug	1	0	1	4
Management science	1	1	0	0
Others	36			
<b>Total</b>	<b>798</b>			

Source: Phạm Duy Hiên, <http://vietnamnet.vn/khoahoc/vande/2006/01/532815/> (1995-2004: 3236)  
 Nguyễn Văn Tuấn: <http://www.tiasang.com.vn/news?id=2166> (1996-2005: 3456)

20

## So sánh Thailand & Vietnam

In 2001-2002

	Thailand	Vietnam	Rate (TL/VN)
# Articles in inter. journals	3103	737	4/1
Made with foreigners (A)	1739	546	3/1
# Citations (A) until 12.2006	13912	4681	3/1
Made in the country (B)	1364	173	8/1
# Citations (B) until 12.2006	5324	323	16/1
Made by universities	1208	69	17/1
# Articles in Math & Physics	68	104	0.7/1

In 2001-2002

	Chulalongkorn	VNU-HN	VNU-HCM	Rate (Chula/ VNUHN+VNUHCM)
# Articles	302	25	8	9.15/1
# Citations	948	63	13	12.47/1
Articles in Math & Physics	15	21	7	0.53/1

Source: Phạm Duy Hiền, <http://vietnamnet.vn/nhandinh/2007/01/649976/>

21

## Đem hội nghị quốc tế đến Vietnam

- Rất ít người từ các nước đang phát triển tham dự được các hội nghị khoa học quốc tế hàng đầu về ICT (NIPS, ICML, KDD, IJCAI, ...)
- Lý do vì không có bài lọt vào các nơi này và không có tiền để đi (thí dụ của IJCAI 2007 tại Ấn độ)
- Hai vấn đề nổi cộm: no show, plagiarism.
- Làm sao đem được nhiều hội nghị quốc tế tốt đến Việt Nam (PAKDD'05, RIVF'07, RIVF'08, PRICAI'08, etc.)?
- Cần sự tham gia với nhiều cố gắng, chuẩn bị và đóng góp từ Việt Nam.



22

## Phần 2

### Xác định đề tài nghiên cứu

(Adapted from the lecture of Prof. Duong Nguyen Vu, HCMC, November 2007)



23

## Finding a research topic: first step

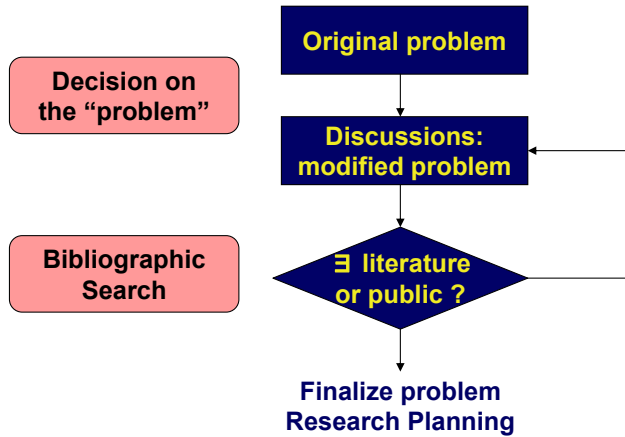
*“The difference between a trivial project and a significant project is not the amount of work required to carry it out, but the amount of thought that you apply in the selection and definition of your problem.”*

David P. Beach & Torsten K.E. Alvager  
*Handbook for Scientific and Technical Research*,  
Prentice-Hall, 1992, p. 29

24

## Phát biểu bài toán

- Quá trình phát biểu bài toán bao hàm một loạt các hành động lặp:



25

## Phát biểu bài toán

- Là bước đầu tiên của mọi đề tài nghiên cứu.
- Nói chung, cái khó nhất của quá trình này là điểm xuất phát: **original idea/topic**.
- Thông thường, chủ đề nghiên cứu được gợi ý bởi thầy hướng dẫn hay phụ trách đề tài: thường được **chọn trong số các bài toán** đang được khảo sát của nhóm hay phòng thí nghiệm.
- Dù thế, ý tưởng khởi đầu thường còn chưa rõ hoặc còn thô → cần phát triển và chế biến.

26

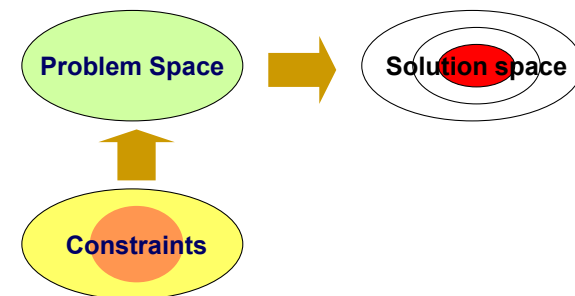
## Nghiên cứu một vấn đề nghiên cứu

- Một **đề tài của luận văn tiến sĩ** hướng đến:
  - Phát triển một lý thuyết mới, một hình thức hóa mới, hoặc
  - Đóng góp vào một lý thuyết hoặc hình thức hóa đã tồn tại.
- Một **đề tài luận văn thạc sĩ** hướng đến:
  - Việc tinh thông các tri thức và kỹ năng của một lĩnh vực trong một nghề (if going to industry)
  - Rèn luyện kinh nghiệm nghiên cứu (if going to doctor course)

27

## Làm mịn một đề tài nghiên cứu

- Các đề tài quá mơ hồ hoặc chưa chính xác đòi hỏi một khối lượng lớn công việc.
- Theo ngôn ngữ toán học, một bài toán có thể có nhiều lời giải được gọi là "*ill-posed problem*."



28

## Hypothesis

- Giả thuyết thể hiện các thành phần của một bài toán nghiên cứu.
- Do vậy, các giả thuyết xác định tập các thí nghiệm, chúng minh cần làm trong quá trình nghiên cứu.
- Trong thực tế, một đề tài nghiên cứu thường chứa đựng nhiều giả thuyết chưa biết.
- Trong quá trình nghiên cứu, người làm nghiên cứu cần làm sáng tỏ các giả thuyết chưa biết này với các minh chứng hoặc chứng minh không thể phủ nhận được.
- Các giả thuyết cần được thiết lập rõ ràng (well-posed.)

29

## Results of a Hypothesis

- Các giả thuyết quy định thí nghiệm cho một lý thuyết – nhằm khẳng định hoặc chỉ ra sự hạn chế của các kết quả đặc biệt được thiết lập từ lý thuyết – cần phải được kiểm chứng theo một trong 4 cách sau:
  - Đối với một phạm vi mở rộng của lý thuyết
  - Đối với các giới hạn của khả năng ứng dụng của lý thuyết
  - Đối với độ chính xác được cải tiến của lý thuyết
  - Đối với sự đánh giá hiệu lực (validation) của các giả thiết cơ bản của lý thuyết.

30

## Hypothesis Validating a Theory

- Kiểm chứng về **phạm vi mở rộng** (extended scope):
  - Mỗi lý thuyết thường chỉ ứng dụng vào được một số tình huống hay điều kiện hạn chế. Lý thuyết này có thể trở nên “hoành tráng” (“powerful”) hơn nếu nó được chứng tỏ có thể áp dụng vào những tình huống khác.
  - Điều ngược lại sẽ củng cố thêm các giới hạn của khả năng ứng dụng của lý thuyết.
- Kiểm chứng về các **giới hạn của khả năng ứng dụng**:
  - Thí dụ: lý thuyết tương đối của Einstein không làm sai lệch cơ học Newton. Lý thuyết này chỉ mô tả các giới hạn trong đó lý thuyết này có thể áp dụng được.

31

## Hypothesis Validating a Theory

- Kiểm chứng việc cải tiến **tính chính xác của lý thuyết**
  - Các lý thuyết thường là sự tổng quát hóa của các hiện tượng quan sát được, qua các độ đo khách quan có được từ sự phân tích trực cảm (heuristics).
  - Tổng quát hóa và tính ứng dụng thường không luôn luôn song hành.
  - Độ chính xác của lý thuyết luôn được mong muốn.
- Kiểm chứng việc **khẳng định hoặc phủ định các giả thiết cơ bản**
  - Liệu giả thiết ranh giới (baseline assumption) có đúng không? Tại sao?
  - Một lý thuyết có thể trở nên kỳ cục (ridiculous) nếu các giả thiết cơ bản không có giá trị khoa học hoặc không có tính thuyết phục với cộng đồng khoa học.
  - Có thể có xung đột khoa học (Scientific “Warfare”!!)

32



## Xác định đề tài nghiên cứu

- Thường có một vài cách để nhìn sâu hơn về đề tài khi bắt đầu làm nghiên cứu:
  - **Phác thảo một tên ngắn gọn** của nghiên cứu: dựa trên giả thuyết ban đầu hoặc “tên công việc”. Chú ý đến sự tiến hóa của giả thuyết hoặc tên công việc.
  - Một chiến lược khác để phát triển chủ đề là đặt cho đề tài một **câu hỏi ngắn** (dạng câu hỏi của giả thuyết).
  - **Tiếp cận Newtonian** – Lời khuyên thực tiễn:
    - Viết ra như “Nghiên cứu của tôi là về ...” ; Tránh dùng một ngôn ngữ phức tạp và uyên bác.
    - Phác thảo ra các tên không dài quá 12 từ, bỏ đi hầu hết các mạo từ (articles) và giới từ (prepositions), và đảm bảo rằng tên này tập trung được vào nội dung chính của nghiên cứu.
    - Định nghĩa các từ trong “giả thuyết” hoặc “tên công việc”
    - Quá trình ssinh nghĩa các từ này cần gắn với khảo sát tài liệu.

33

## Khảo sát tài liệu

- Việc khảo sát tài liệu giúp người nghiên cứu **thu hẹp phạm vi** của việc tìm hiểu và chuyển tầm quan trọng của việc tìm hiểu đề tài tới người đọc.
- Khảo sát tài là **một phần không thể thiếu** của việc lựa chọn/quyết định đề tài nghiên cứu.
  - Cần xem xét liệu đề tài *có thể và cần phải* khảo sát.
  - Phải xác định được một đề tài ban đầu (bởi người làm nghiên cứu thay vì người hướng dẫn).
  - Khảo sát tài liệu liên quan việc “đối thoại” không ngừng với tư liệu về đề tài, lấp những chỗ trống về hiểu biết và mở rộng những điều đã biết.
  - Cung cấp một sườn làm việc để thiết lập tầm quan trọng của nghiên cứu cũng như một chuẩn để so sánh các kết quả đã có với những phát kiến mới.

34

## Tìm kiếm tư liệu

- Tìm kiếm tài liệu cho phép biết những tri thức về lĩnh vực (domain knowledge), về các nghiên cứu liên quan để vẽ lên được một bức tranh theo thời gian các nghiên cứu về chủ đề.
- Cần nhận rõ là yếu tố cốt tử của nghiên cứu là tri thức tìm ra được **phải mới hoặc đầu tiên** (**must be new or original.**)
- Một nghiên cứu chỉ tìm ra kết quả đã được cộng đồng biết từ trước, và chỉ có ích cho một người hoặc một nhóm nhỏ, sẽ **không có giá trị như một nghiên cứu.**
- Có những nghiên cứu đã bị gián đoạn trong những bước đầu bởi nhiều lý do khác nhau, nhưng phần lớn còn lại đã bị ngừng vì đề tài nghiên cứu không đủ “thuyết phục”. Cũng không hẳn là bất thường nếu tiếp tục triển khai đề tài này sau việc khảo sát kỹ tư liệu.

35

## Các bước khảo sát tư liệu

*“A literature review for a proposal or a research study means locating and summarizing the studies about the topic.”*

- Step 1: Bắt đầu bằng việc xác định các **keywords** cần thiết để tìm kiếm tài liệu. Các từ khóa này có thể nảy sinh khi xác định một đề tài.
- Step 2: Tìm kiếm tài liệu trong các cơ sở dữ liệu của thư viện (**library databases**) với các từ khóa đã xác định.
- Step 3: Thử tìm khoảng **50 reports and articles** (or books) liên quan đến chủ đề nghiên cứu. Sắp xếp các tư liệu này theo thứ tự ưu tiên.
- Step 4: Đọc nhanh các tài liệu chọn lọc với tập trung về các **abstracts and core results** để thu được cảm nhận liệu các tài liệu này có đem tới một đóng góp có ích cho việc hiểu đề tài.

36

## Các bước khảo sát tư liệu

- Step 5: Với mỗi tài liệu được chọn, viết ra **maximum 10 lines** để mô tả bạn đã hiểu tài liệu này thế nào, tức việc hiểu về đóng góp của tài liệu đối với chủ đề nghiên cứu.
- Step 6: Thiết lập một **"literature map,"** tức một bức tranh về các tư liệu nghiên cứu về chủ đề. Bức tranh này cung cấp một phương tiện để đặt nghiên cứu của bạn vào trong một toàn thể lớn hơn các tư liệu về chủ đề.
- Step 7: Cùng lúc với (6), bắt đầu phác thảo các **summaries** của các bài báo liên quan nhất khi dùng các kết quả của bước (5). Các tóm tắt này được kết hợp vào các khảo sát tư liệu cuối cho đề xuất nghiên cứu. Chú ý để đảm bảo có các tài liệu tham khảo chính xác.
- Step 8: Thảo luận với **peers and supervisors** về nghiên cứu của bạn; hoặc lặp quá trình này hoặc cuối cùng "lắp ráp" các tài liệu được khảo sát sao cho chúng được tổ chức hoặc cấu trúc bởi các khái niệm đã được đề cập. Kết thúc với một tóm tắt các chủ đề chính tìm được trước khi gợi ý xa hơn các chủ đề nghiên cứu cần thiết.

37

## Using the Internet

- Internet không phải là một thư viện
- Books and journals trong một thư viện đều có người chịu trách nhiệm và nhà xuất bản.
- Bất kỳ ai cũng có thể đưa bất kỳ thứ gì lên Internet.
- Cần chú ý đến sự đáng tin của nguồn tài liệu (cấp bởi các viện, các tổ chức tư nhân có uy tín, các cơ quan chính phủ, etc).
- Cần phân biệt phân biệt cái hay cái dở trên Internet (to separate the wheat from the chaff).

38

## Hướng dẫn đánh giá thông tin (1)

Ref. Elizabeth Kirk (2002) - Johns Hopkins University Library Web page  
[www.library.jhu.edu/elp/useit/evaluate/index.html](http://www.library.jhu.edu/elp/useit/evaluate/index.html)

### NGUỒN GỐC TÁC GIẢ (AUTHORSHIP)

- Tác giả có là người nổi tiếng?
  - Nếu không, công trình của tác giả này có được trích dẫn bởi các tác giả nổi tiếng khác?
    - Nếu không, liệu bạn có thể tìm thấy thông tin tư liệu để sự tin cậy của tác giả?

### NGUỒN GỐC NƠI XUẤT BẢN (PUBLISHING BODY)

- Ai là người tài trợ cho web site?
- Trang Web có phải của một tổ chức có uy tín? Thông tin có trên các trang Web cá nhân cần phải xem như đáng ngờ (considered highly suspect.)

### QUAN ĐIỂM (POINT OF VIEW)

- Liệu tổ chức tài trợ trang Web có một quan điểm?  
Note: Cần xem thông tin của các công ty kinh doanh gắn với quảng cáo  
Cần đặc biệt cảnh giác với thông tin cung cấp bởi các nhóm luật sư (advocacy groups.)

39

## Hướng dẫn đánh giá thông tin(2)

### LIÊN KẾT TƯ LIỆU (CONNECTION TO THE LITERATURE)

- Có các tài liệu tham khảo tới các công trình trong lĩnh vực không?
- Có các lý thuyết thích hợp được thảo luận?
- Các tranh luận có được thừa nhận?

### TÍNH THẨM TRA ĐƯỢC (VERIFIABILITY)

- Có thông tin cho phép bạn kiểm chứng phương pháp?

### TÍNH THỜI GIAN (CURRENCY)

- Có không một ngày trên tài liệu cho phép bạn đánh giá về thời gian của thông tin?

### SEARCH ENGINE

- Search engine đã xác định thế nào thứ tự các hits?
- Some search engines bán chỗ cho nhà quảng cáo. Bạn đã thấy search engine xác định thứ tự thông tin tìm được thế nào?

40

## Đánh giá tính khả thi (1)

- Nếu cơ may của thành công và thất bại của một khảo sát có thể tiên liệu được (anticipated) ở một mức khá chính xác trong giai đoạn xác định nghiên cứu, ta có thể tránh được các nỗ lực vô ích.
- Tính khả thi của một vấn đề nghiên cứu có thể được đánh giá dưới những góc độ sau:
  - **Lĩnh vực khảo sát:** Đề tài nghiên cứu là phát triển của một quá trình suy nghĩ, một ý tưởng, hoặc của cả hai? Đây là các lĩnh vực khoa học của đề tài? Bạn có đủ tri thức để đánh giá đề tài?
  - **Vấn đề nghiên cứu:** Liệu chúng ta có thể minh học chủ đề nghiên cứu bởi một mô tả đơn giản dưới dạng một câu hỏi?

41

## Đánh giá tính khả thi (2)

- **Tính sẵn sàng của dữ liệu cơ bản:** Liệu bạn có thể có những dữ liệu cơ bản cần thiết?
  - Quen thuộc với dữ liệu thô,
  - Nguồn dữ liệu,
  - Phương pháp thu thập dữ liệu,
  - Yêu cầu cho các trang bị đặc biệt,
  - Thao tác trên các thiết bị
- **Đánh giá khả năng:** đây là lĩnh vực khoa học và công nghệ thứ nhất, thứ hai cần thiết cho việc tìm kiếm giải pháp? Liệu ta có khả năng tiến hành nghiên cứu trong các lĩnh vực này?

42

## Tiêu chuẩn quyết định

- **Tiêu chuẩn chính:**
  - **Ham thích cá nhân:** đề tài nghiên cứu phải kích thích trí tưởng tượng và sáng tạo. Người nghiên cứu cần phải “be in.”
  - **Kích cỡ của nghiên cứu:** đề tài nghiên cứu cần phải quản lý được.
  - **Khả năng của nhóm nghiên cứu**
  - **Triển vọng** của tính mới mẻ
  - **Thông thạo** các lĩnh vực liên quan chủ đề nghiên cứu.

43

## Tiêu chuẩn quyết định

- **Các tiêu chuẩn khác:**
  - **Timing :** Đề tài có “hot” không?
  - **Originality:** Đã có ai làm chưa?
  - **Solidity:** Đề xuất nghiên cứu đã đầy đủ chưa? Có mâu thuẫn gì không?
  - **Utility:** Phạm vi của nghiên cứu là gì? Có ích lợi không? Dùng cho ai? Các lĩnh vực ứng dụng là gì?
  - **Morality:** Có vấn đề gì về đạo đức không?
  - **Feasibility:** Những ràng buộc đi cùng là gì?
    - *Human Resources? Time?*
    - *Financial Resources? Cost?*
    - *Availability of initial data?*

44

## Phần 3

### Trình bày và giới thiệu kết quả nghiên cứu: Papers, Talks, and Chats

(from tutorial of Prof. Marie desJardins (University of Maryland)  
ICML/KDD 2003

45

## Research isn't just research

- Ai quan tâm đến việc bạn làm nếu bạn không bao giờ nói cho họ biết?
- Bạn cần phải giới thiệu ý tưởng của mình ở nhiều dạng khác nhau:
  - Liên kết mạng với các bạn bè
  - **Viết và gửi bài tới các** workshops, conferences, and journals
  - **Giới thiệu các bài báo** ở workshops and conferences
  - Đặt chúng lên một website để có thể nhấn mạnh điều bạn quan tâm và hoạt động nghiên cứu

46

## Networking

- **Meet people!** Giúp cho có một mục tiêu:
  - Xem xem họ đang làm nghiên cứu gì
  - Nói với họ bạn đang làm gì
  - Tìm một lĩnh vực có chung quan tâm
  - Học cách nhìn nhận và các phương hướng tương lai của họ
  - Thảo luận một hướng nghiên cứu mới hoặc đề tài mới
- Bạn thu được gì từ giao tiếp này?
- Họ thu được gì?
- Nếu bạn biết hai người bạn, và họ biết hai người bạn khác, và những người bạn này biết hai người khác, ... Bạn sẽ sớm quen biết rất nhiều người!

47

## Networking II

- Bạn cần chuẩn bị để **tóm tắt được nghiên cứu của mình**
  - Cần sẵn sàng giới thiệu chủ đề luận án trong vòng 1 phút, 5 phút, và 15 phút
  - Tương tự với các loại đề tài nghiên cứu khác
  - Cần biết phân biệt rõ đóng góp của riêng bạn, của thầy hướng dẫn, và ý tưởng từ các nghiên cứu trước đó.
  - Thực hành với các sinh viên khác!

48

## Writing and submitting papers

- Đối với **master's thesis**, cần phấn đấu để có ít nhất một bài báo ở hội nghị “good” conference vào lúc tốt nghiệp
- Đối với **doctoral dissertation**, cần phấn đấu để có vài bài hội nghị tốt và một bài tạp chí
- Viết những bài báo này là một thực hành lớn cho chính luận văn của mình ... (và bạn có thể dùng lại các vật liệu này!)
- Nộp bài ở đâu?
  - Nhìn danh sách các ấn phẩm của những người làm nghiên cứu liên quan, và đọc công bố của họ
  - Công bố ở các hội nghị có những bài hay nhất

49

## Writing papers: Strategy

- Trước hết, quyết định xem nộp bài vào đâu
  - Bạn có thể không kịp nộp bài đúng hạn, nhưng việc có một deadline là luôn cần thiết
  - Hai đến bốn tháng luôn là thời hạn tốt để kế hoạch
- Tiếp theo, quyết định xem nói gì
  - Ý tưởng chính là gì? Đã phát triển chúng chưa?
  - Kết quả chính là gì? Đã thiết kế và chạy các thí nghiệm chưa? Đã phân tích dữ liệu chưa?
  - Các nghiên cứu liên quan chính là gì? Bạn đã có và đọc các bài này chưa? Liệu bạn có thể tóm tắt chúng?
- Now get started on the work you need to do to fill in the missing holes! (You can write in parallel...)

50

## Writing papers: Design

- **Abstract** – tóm tắt **research contributions**, không phải tóm tắt bài báo (i.e., không thể là dàn ý của bài báo)
- **Introduction/motivation** – **những gì bạn đã làm** và **tại sao người đọc cần quan tâm**, cộng thêm dàn ý của bài báo
- **Technical sections** – một hoặc nhiều sections tóm tắt và giới thiệu các ý tưởng nghiên cứu bạn đã phát triển
- **Experiments/results/analysis** – một hoặc nhiều sections giới thiệu kết quả thực nghiệm
- **Future work** – tóm tắt những gì bạn dự định làm và những câu hỏi mở còn cần phải trả lời.
- **Conclusions** – nhắc lại những gì đã làm và tại sao chúng quan trọng
- **Related work** – đôi khi viết ngay sau phần introduction, đôi khi viết trước phần conclusions (phụ thuộc chừng mực nào vào liệu việc bạn làm dựa trên nghiên cứu trước kia, hoặc phủ nhận chúng)

51

## Writing papers: Tactics

- **Top-down design** (outline) rất quan trọng
- Các gạch đầu dòng có thể giúp bạn nhận rõ cấu trúc bài viết
  - Trừ khi bạn là người rất tài hay rất nhiều kinh nghiệm, bạn cần thiết kế cẩn thận *trước khi* bắt tay vào viết bài
- Neatness counts! Kiểm tra spelling, grammar, consistency of fonts and notation *before* chúng đến với mắt người khác
  - Nếu người đánh giá rất chú ý đến các lỗi typos của bạn, họ có thể bỏ sót những gì hay của nội dung
- Dành thời gian để kiểm tra!
  - Fellow students, collaborators, advisors, ...
  - Một bài báo được viết khi gửi bài ... Nhưng thường cũng không hẳn chỉ khi này.

52

## Authorship

- Who should be an author?
  - Những người có đóng góp đáng kể vào việc phát triển các nội dung bài báo hoặc viết bài báo
  - Không *nhất thiết cần* có tên người đọc góp ý, mã hóa, chạy thí nghiệm
- What order should the authors be listed in?
  - Nếu có những tác giả đóng góp nhiều hơn người khác về việc phát triển nội dung và/hoặc viết phần lớn hay tất cả bài báo, họ phải được kể lên trước
  - Nếu đóng góp là như nhau hoặc tác giả cùng việc trong một nhóm, tên nên viết theo thứ tự ABC
  - Đôi khi nên có thêm chú thích “The authors are listed in alphabetical order”

53

## Giving talks

- **Cần biết bạn được nói bao lâu**
  - Thời gian nói là bao lâu? QA có bao gồm trong thời gian đó?
  - Sẽ rất tốt nếu mỗi slide được nói trong vòng 2-3 phút
  - Nếu bạn có quá nhiều slides, bạn phải bỏ qua một số, thậm chí dở hơn là phải bỏ hết phần còn lại vì hết giờ. Hãy tránh các tình huống này!!
  - Bạn không bao giờ có thời gian để nói mọi thứ về đề tài, do vậy đừng băn khoăn nếu phải bỏ qua một số nội dung!
  - Trừ khi bạn là người nói rất nhiều kinh nghiệm, bạn cần phải thực hành với việc tính đến thời gian:
    - A couple of times on your own to get the general flow
    - At least one dry run to work out the kinks
    - A run-through on your own the night before the talk

54

## Giving talks II

- **Cần biết người nghe là ai**
  - Đừng phí thời gian về những thứ cơ bản nếu bạn đang nói với những người đang làm cùng nghề
  - Thậm chí với những người này, cần chắc là bạn giải thích những khái niệm mới một cách rõ ràng
  - Mặt khác, bạn sẽ mất người nghe không cùng làm nghề nếu không cung cấp những kiến thức cơ bản
  - Trong mọi trường hợp, điều quan trọng nhất là cần nhấn mạnh what you've done và why they should care!

55

## Giving talks III

- **Biết rõ bạn muốn nói gì**
  - Chỉ nêu ra một tóm tắt của đề tài sẽ chẳng có gì thích thú với hầu hết người nghe.
  - Bạn cần nêu đủ chi tiết để người nghe chú ý đến ý tưởng của bạn (và chỉ ra bạn đã thực làm gì, nhưng không cần quá chi tiết)
  - Họ muốn nghe what you did that was cool và why they should care
  - Thông thường, họ thích nghe hai điều trên khi bắt đầu, trong quá trình, và ở phần cuối của trình bày.
  - Nếu người nghe thích, họ sẽ hỏi hoặc đọc bài báo
  - Trong mọi trường hợp, không được chỉ đọc các slides!  
→ Báo cáo hội nghị là “kể một câu chuyện”

56

## Preparing slides

- Don't just read your slides!
- Use the minimum amount of text necessary
- Use examples
- Use a readable, simple, yet elegant format
- Use color to emphasize important points, but avoid the excessive use of color
- “Hiding” bullets like this is annoying (but sometimes effective), but...  
Abuse of animation is a cardinal sin!
- Don't fidget, and...
- Don't just read your slides!

57

## How to give a bad talk

Advice from Dave Patterson, summarized by Mark Hill

1. Thou shalt not be neat (rõ ràng)
2. Thou shalt not waste space
3. Thou shalt not covet brevity (muốn sự ngắn gọn)
4. Thou shalt cover thy naked slides (trơ trụi)
5. Thou shalt not write large
6. Thou shalt not use color
7. Thou shalt not illustrate
8. Thou shalt not make eye contact
9. Thou shalt not skip slides in a long talk
10. Thou shalt not practice

58

## Some useful resources

- Writing:
  - Lynn DuPre, *Bugs in Writing*
  - Strunk & White, *Elements of Style*
- Giving talks:
  - Mark Hill, “Oral presentation advice”
  - Patrick Winston, “Some lecturing heuristics”
  - Simon L. Peyton Jones et al., “How to give a good research talk”
  - Dave Patterson, “How to have a bad career in research/academia”

59

## Phần 4

### Một vài kinh nghiệm cá nhân

(from what I have discussed with students from Vietnam)

60

## English pronunciation

- Usually do not pronounce vowels
- Examples
  - book vs. books
  - Text mining vs “tech” mining
  - Adapter [a-da-pu-ta]
  - She asks me if I can fly to the moon

61

## Research proposal (typically in Japan)

- **Objectives**
  - What to do clearly
  - Its significance
- **Background and research context**
  - Who are doing similar research, related research?
  - What are approaches to solve the problems?
  - Your critical view on the related work
- **Methodology and plan**
  - What could be the key idea of the solution?

62

## If I may give an advice to students

- To be self-confident



63