

# Deja vu

## กระบวนการพื้นฐานของ Deja vu

บทความที่มีความเหมือนกันสูง พร้อมที่จะนำเข้าสู่ Déjà vu มาจาก 3 ช่องทาง คือ

- **eTBLAST** แต่เดิมทีมงานเลือกสุ่มตัวอย่างบทความจาก PubMed นำมาประมวลผลด้วย eTBLAST โดยใช้ Title และ Abstract เป็น Key word เพื่อหาความเหมือนกันกับบทความอื่น ซึ่งต่อมาพบว่า 75% ของคู่บทความที่เหมือนกัน อยู่ในส่วนที่ฐานข้อมูล PubMed จัดไว้เป็น “Related Articles” ดังนั้น กระบวนการจึงปรับไปใช้การเลือกบทความที่ใกล้เคียงที่สุดของ Related Articles (The topmost “Related Articles”) มาประมวลผลโดย eTBLAST เทียบกับบทความข้างเคียงของตัวเอง ด้วยวิธีการเช่นนี้ ทำให้ทีมงานสามารถประหยัดเวลา จากการสุ่มบทความของ PubMed ที่มากกว่า 17 ล้านบทความ และแต่ละบทความใช้เวลาประมวลผลประมาณ 40 วินาที ไปได้อย่างมาก ซึ่งหากวิเคราะห์แบบเดิม จะต้องใช้เวลาราว 10 ปี เหลือเพียง 10 วัน และได้บทความที่มีความเหมือน จำนวน 74,777 บทความ ซึ่งนำเข้าไปในฐานข้อมูล เพื่อรอการวิเคราะห์ตามขั้นตอนต่อไปโดยทีมงานเรียบร้อยแล้ว
- **Manual detection**
- Submitted (form users) ตามแบบฟอร์ม ดังรูปที่ 1

**Report a Duplicate Citation**

This section allows users to:

- Report a case of suspected duplication (with similar or unrelated authors), then our curators will evaluate the submission and decide to add it or not Deja Vu.
- Contest a record currently in Deja Vu. A justification is required and supplemental data may be requested. Actions in response to a contest may span from immediate re-classification of a duplicate within the database to the addition of the comment to the database in response to subjective differences.

In all cases we will ask you to provide us with your contact information.

First name:  Last name:

email:

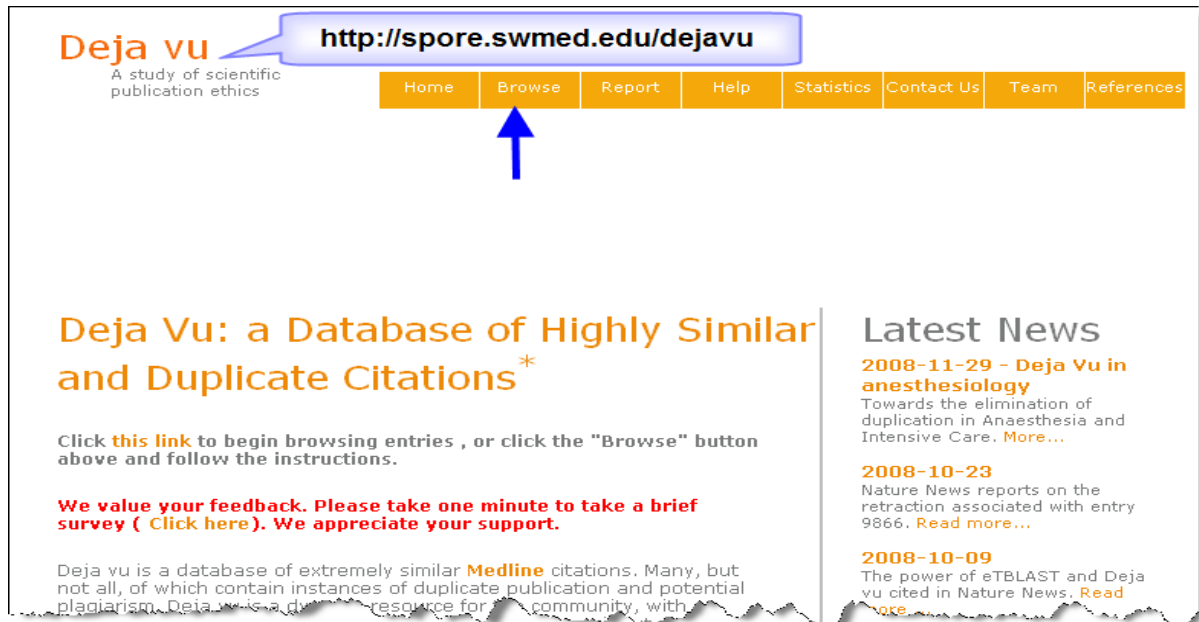
Action:

Below, please provide specific information: links to Pubmed records or PMIDS you wish to report as duplicate citation, or link to Deja vu record (s) you wish to contest. Also a justification is required explaining the reasons for your action (few lines). If no justification is provided, this report may not be considered.

รูปที่ 1

## Déjà vu Searching

- ที่ URL : <http://spore.swmed.edu/dejavu> คือหน้า Home ของฐานข้อมูล คลิกเลือกที่ Browse (รูปที่ 2)



รูปที่ 2

- ฐานข้อมูลประกอบด้วยข้อมูลใน 2 ตาราง คือ Entries และ Articles สำหรับบทความที่ผ่านกระบวนการของ eTBLAST และ การวิเคราะห์จะอยู่ในตาราง Entries (รูปที่ 3)

**Click on the "Entries" link below to access the data.**

Currently, there are two tables available for browsing and searching. Most users want to search the list of paired highly similar entries. Once you have clicked on the entries link below, the list of entries will be presented on a new page. On that page, there are several ways to best navigate and inspect the paired highly similar entry list. First, any pair of highly similar citations can be viewed side-by-side by clicking on the ID number (far left column). Second, the table can be filtered to inspect a list of only a particular type of entries by clicking on one of the filters to the far right. Finally, the table can be searched for any keyword (an author, institution, etc.) using the box in the upper left and then clicking on the "Go" button.

<b>1</b> Entries	Repository of curated and computationally identified highly similar citation pairs.
<b>2</b> Articles	List of articles (pmid/database/reference) contained in Deja vu.

รูปที่ 3

- ในตาราง Entries ประกอบด้วยข้อมูลของบทความคู่ที่เหมือนกัน จำนวน 74,777 บทความ พร้อมรายละเอียดข้อมูลจากการวิเคราะห์ และช่องว่างสำหรับใส่คำค้นเมื่อต้องการสืบค้นข้อมูล (รูปที่ 4)

1

รูปที่ 4

- 1 **Query Box** ช่องว่างสำหรับใส่คำค้นเมื่อต้องการสืบค้นข้อมูล โดยคำค้นอาจเป็น ชื่อเรื่อง (Title) ชื่อผู้แต่ง (Author) ชื่อวารสาร (Journal) ก็ได้
- 2 **Detailed Search** ใช้สำหรับสืบค้น เมื่อต้องการค้นหาด้วยคำค้นหลายแบบ ซึ่งมีให้เลือกใช้ ดังนี้ (รูปที่ 5)

**Search Fields**

**Which:** search in  ▼

**PMID:** exactly matches

**Journal:** exactly matches  [Journal list](#)

**Author:** contains  (Case sensitive)

**Address:** contains

**Title:** contains

**Year:** published in

รูปที่ 5

3

รายละเอียดข้อมูลของแต่ละบทความ คือ Earlier Article, Later Article, Lag (ระยะเวลาเป็นเดือน ระหว่างการตีพิมพ์ของบทความแรกและบทความหลัง) , Language, Identity, Sim. Score, Ratio, Share author, Classification, Modified date

สำหรับ Classification ของบทความใน Déjà vu แบ่งออกเป็น 6 แบบ โดยมีจำนวนบทความในแต่ละแบบ ดังนี้ (รูปที่ 6)

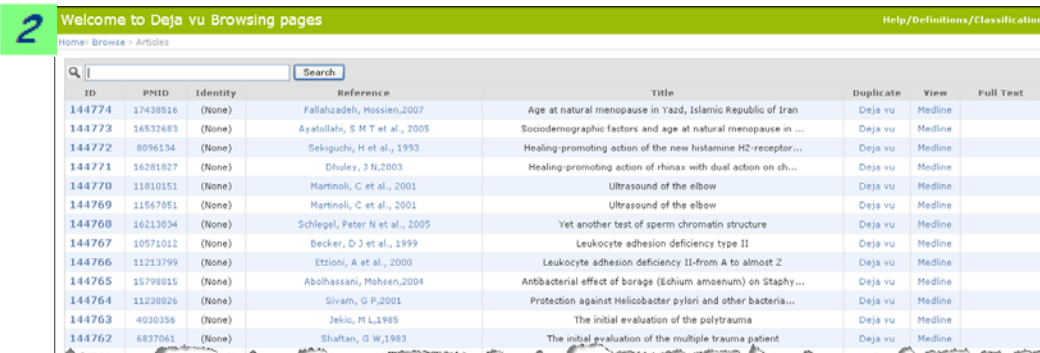
Entry type	Count
DISTINCT	1452
DUPLICATE	2098
ERRATUM	191
NO ABSTRACT	16
SANCTIONED	1689
UNVERIFIED	68850
UPDATE	481
<b>TOTAL</b>	<b>74777</b>

รูปที่ 6

4

Display options สำหรับใช้เป็นทางเลือกในการแสดงผล โดยสามารถกำหนดเงื่อนไขตามที่ต้องการได้ เช่น

- Duplicate SA (Duplicate with shared authors)
- Duplicate DA (Duplicate with no shared authors)
- DISTINCT (Display only entries classified as distinct and unique work)
- DUPLICATE (Display only entries classified as duplicates)
- UNVERIFIED (Display only entries that have yet to be classified)
- Journals (SJ/DJ) : Display only entries that have same Jr. or different Jr. เป็นต้น

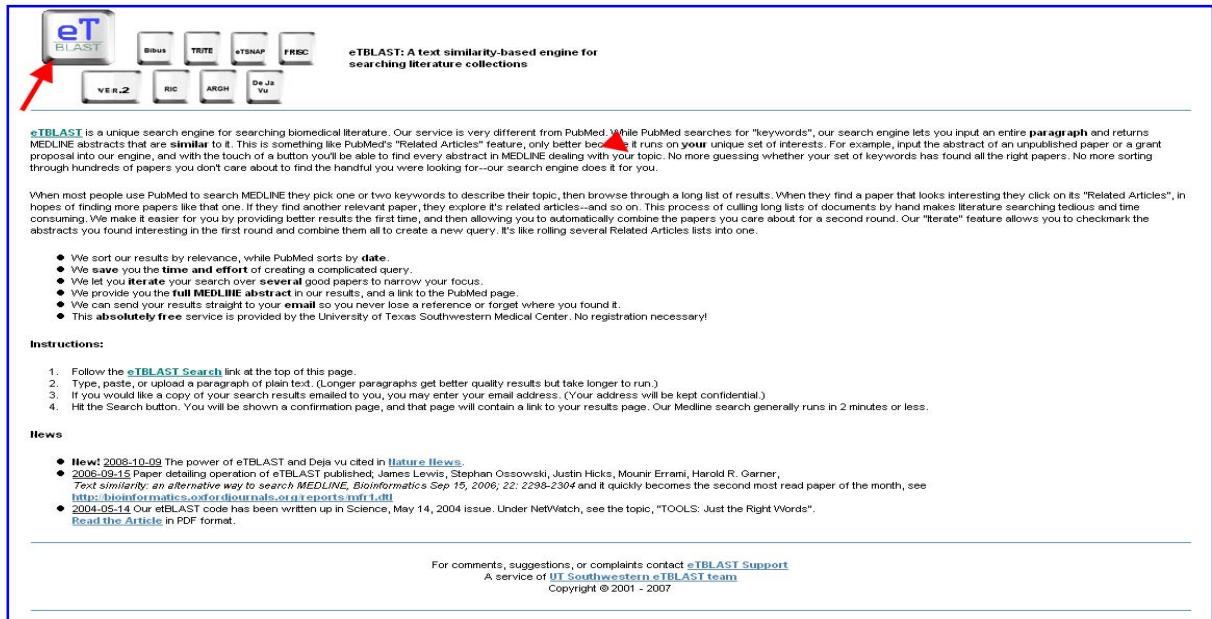


รูปที่ 7

ตาราง Articles สำหรับใช้ข้อมูล pmid , database และ reference ของบทความในฐานข้อมูล Déjà vu

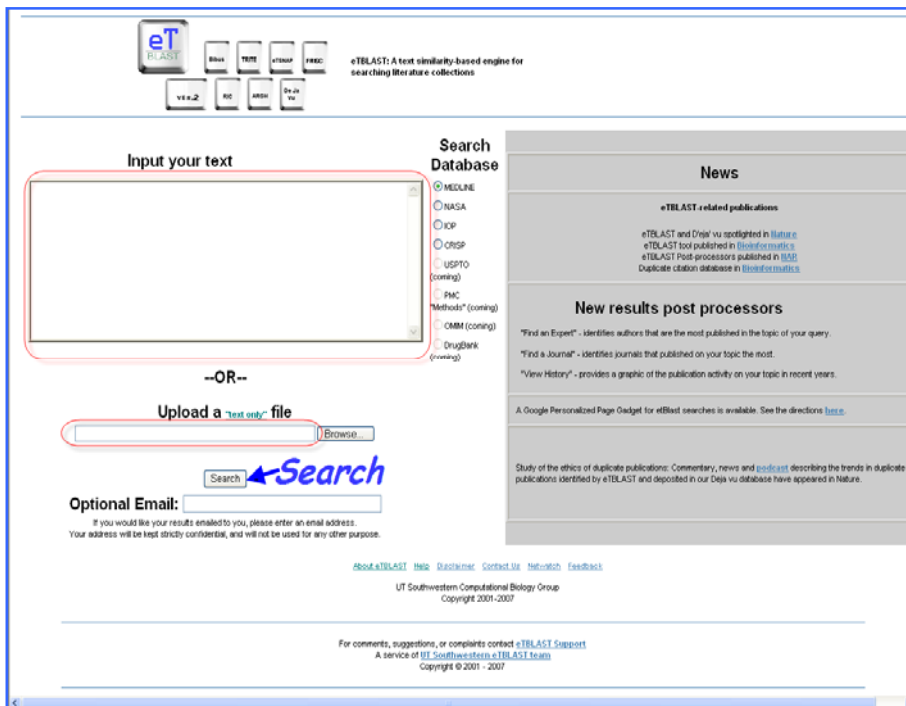
### eTBLAST Searching

- ที่ URL : <http://invention.swmed.edu/etblast> คือ หน้า Homepage ของโปรแกรม คลินิกที่ eTBALST (รูปที่ 8)



รูปที่ 8

- ใส่คำค้นลงในช่อง Input your text ด้วย Title หรือ Author หรือ Abstract หรือในกรณีมี Text file อยู่แล้ว สามารถ Upload ขึ้นมาสืบค้นได้ทันที โดยใช้ช่อง Upload a “text only” file จากนั้น คลิก Search (รูปที่ 9)



รูปที่ 9

- eTBALST ใช้เวลาในการประมวลผลครั้งหนึ่งๆประมาณ 1-2 นาที (รูปที่ 10)

eTBLAST: A text similarity-based engine for searching literature collections

**Submission Successful!**  
Please do not reload this page, your results are being computed.

This page checks for your results every 10 seconds until your results are ready.

seconds until we check for your results.  
Typical times for retrieving results are from 1 to 2 minutes.

Your eTBLAST results will be available shortly and will be available at <http://www.ewmed.edu/cgi-bin/etblast/etblast.pl?user-id=329779545> The time it takes to process your submission depends on the size of your submission.

รูปที่ 10

ผลการประมวลผลโดย eTBLAST (รูปที่ 11)

eTBLAST: A text similarity-based engine for searching literature collections

**Your submission self score is: 231.525**

**Most Similar Matches to Your Query:**

- [\[Osteoclastic resorption of Haversian system of femoral neck cortex in aged women\]](#) **highly similar match!**  
B Chai ... Z Tan  
Zhonghua Wai Ke Za Zhi 1995 May; 33(5):313-5. Score: 223.901 Z-score: 1470.5
- [Osteoclastic resorption of Haversian systems in cortical bone of femoral neck in aged women. A scanning electron microscopic study.](#) **highly similar match!**  
B Chai ... H Li  
Chin Med J (Engl) 1996 Sep; 109(9):705-10. Score: 187.999 Z-score: 1234.302
- [Histopathologic study of bone invasion by DMBA-induced carcinoma of the mouth in the hamster.](#)  
Y Totsuka ... K Tomita  
Oral Surg Oral Med Oral Pathol 1986 Dec; 62(6):683-92. Score: 21.709 Z-score: 140.292
- [Microradiographic and histomorphometric indices of mandibles for diagnosis of osteopenia.](#)  
Mahn Wowerp ...  
Score: 24.153 Z-score: 136.83

รูปที่ 11

โปรแกรมจะแสดงค่าต่างๆ ที่ใช้บ่งชี้ความเหมือน เช่น Score , self score , Z-score ตามลำดับจากมากไปน้อยของบทความที่พบ นอกจากนี้โปรแกรมยังให้ข้อมูลเพิ่มเติมเกี่ยวกับเรื่องที่ใช้สืบค้น เช่น

- Visualize Pair wise Comparisons
- Find an Expert in this Field
- Find a Journal for your Manuscript
- View the Publication History of this Topic

ชมพูนุช บุญญวรรณ

ห้องสมุดคณะแพทยศาสตร์

มหาวิทยาลัยเชียงใหม่

ธันวาคม 2551

## References

1. Déjà vu: a Database of Highly Similar and Duplicate Citations. Retrieved December 15, 2008, from <http://spore.swmed.edu/dejavu>
2. eTBLAST: A text similarity-based engine for searching literature collections. Retrieved December 15, 2008, from <http://invention.swmed.edu/etblast>
3. Errami M., Garner H. A tale of two citations. Nature commentary 451/24, 397-399(2008)
4. Butler D. Entire-paper plagiarism caught by software : Thousands of “similarities” found between papers. Nature news. Publish online, doi:10.1038/455715a (2008).
5. กิตติศักดิ์ กุลวิจิต วรรศักดิ์ โชติเลอศักดิ์ วลี ตูวรรณนะ “PLAGIARISM ในการเขียนบทความวิจัย” : การอบรม , สำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่ 11 พฤศจิกายน 2551.