การประยุกต์ใช้ AI ในงานเทคนิค

เอกสารประกอบการเสวนาการประชุมวิชาการ
“AI พันธมิตรใหม่ของบรรณารักษ์”
จัดโดย ชมรมบรรณารักษ์สถาบันอุดมศึกษา
10 ก.พ. 2568 เวลา 8.00-12.00 น.

นายรัฐธีร์ ปภัสสุรีย์โชติ
สำนักงานวิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การใช้ AI ในงานเทคนิค

- การจัดหาทรัพยากรสารสนเทศ
    - Gap analysis
    - วิเคราะห์ข้อมูลการใช้ทรัพยากรสารสนเทศของผู้รับบริการ แล้วใช้ในการคาดเดาหรือพยากรณ์ว่าควรจัดหาทรัพยากรสารสนเทศที่มีเนื้อหาลักษณะใดเข้ามาให้บริการ หรือใช้ในการคัดทรัพยากรสารสนเทศออก
    - วิเคราะห์ข้อมูล ดู trend ของเนื้อหาที่น่าจะเป็นที่สนใจต่อไปในอนาคตเพื่อใช้เป็นข้อมูลในการจัดหาทรัพยากรสารสนเทศ

การใช้ AI ในงานเทคนิค

- การวิเคราะห์ทรัพยากรสารสนเทศ
  - จัดทำ metadata
  - ช่วยเพิ่มความสมบูรณ์ของ metadata

ป้อนข้อมูลโดยตรง หรือนำไฟล์หนังสือหรือทรัพยากรสารสนเทศเข้าไปให้ AI อ่าน

- จัดทำ metadata และจัดระบบ เช่น ให้หัวเรื่องและเลขหมู่
- ช่วยทำสาระสังเขปหรือเรื่องย่อ

การสร้าง MyGPT สำหรับงานวิเคราะห์ทรัพยากรสารสนเทศ: ประสบการณ์ของสำนักงานวิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

นำเสนอผลงานด้วยการบรรยาย
กลุ่มการจัดการทรัพยากรสารสนเทศ
ประชุมวิชาการระดับชาติ PULINET ครั้งที่ 15
วันที่ 16 มกราคม 2568

นายรัฐธีร์ ปภัสสุรีย์โชติ
นายอภิวัฒน์ แก้วหะวงษ์
สำนักงานวิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

วัตถุประสงค์

ห้องสมุดคณะ วิทยาลัย สถาบัน
ขาดบุคลากรด้านงานวิเคราะห์ทรัพยากรสารสนเทศ

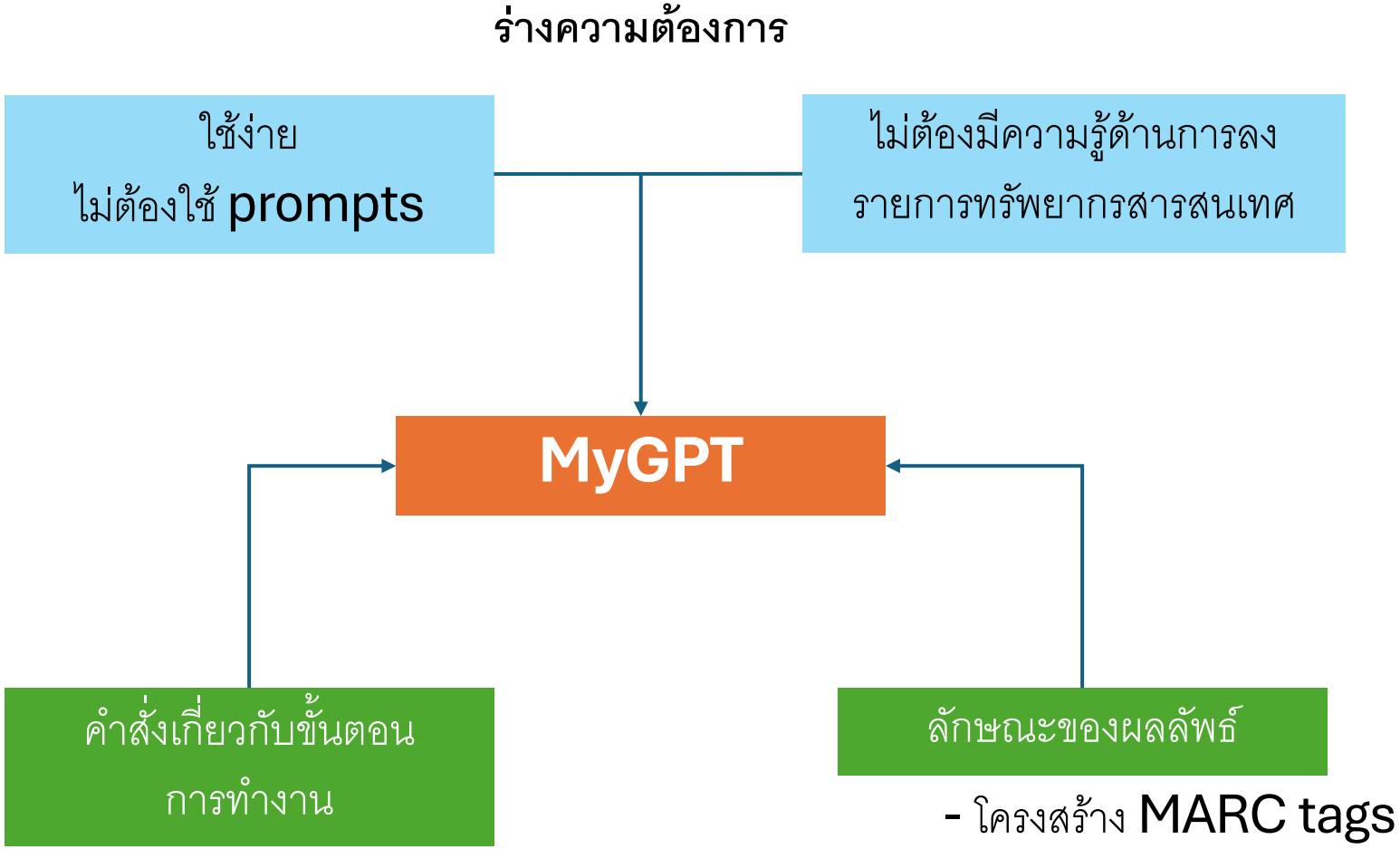
ปัญญาประดิษฐ์เริ่มเป็นที่นิยมใช้งาน
อย่างแพร่หลาย

สามารถนำ AI มาช่วย
หรือใช้กับงานวิเคราะห์ทรัพยากรสารสนเทศได้หรือไม่ อย่างไร

พิจารณาจาก

- **วิธีการ**
- **ความถูกต้องของข้อมูล**
- **รูปแบบความผิดพลาด**

วิธีการดำเนินการ



เขตข้อมูล MARC	Leader		008		020		040		336, 337, 338		082		เลขผู้แต่ง		100		245		264		505		520		650		7XX		ผลรวม	
คะแนน	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#	B**	C#
จำนวนครั้ง ที่ถูกต้องระดับ 2	1	0	1	2	10	10	8	10	10	10	5	3	1	0	8	8	3	10	3	0	3	10	0	10	0	1	2	9	55	83
จำนวนครั้ง ที่ถูกต้องระดับ 1	7	10	0	0	0	0	2	0	0	0	1	6	0	0	0	2	1	0	1	0	0	0	9	0	0	4	0	1	21	23
คะแนนรวม ด้านความถูกต้อง	9	10	2	4	20	20	18	20	20	20	11	12	2	0	16	18	7	20	7	0	6	20	9	20	0	6	4	19	131	189
คะแนน ด้านโครงสร้าง	10	10	9	10	9	10	9	10	10	10	7	10	5	10	8	10	8	3	0	10	0	10	10	10	2	1	8	10	95	124
คะแนนรวมทั้งหมด***	19	20	11	14	29	30	27	30	30	30	18	22	7	10	24	28	15	23	7	10	6	30	19	30	2	7	12	29	226	313
ร้อยละ	*64*	*67*	*37*	*47*	*97*	*100*	*90*	*100*	*100*	*100*	*60*	*74*	*24*	*34*	*80*	*94*	*50*	*77*	*24*	*34*	*20*	*100*	*64*	*100*	*7*	*24*	*40*	*97*	*54*	*75*
เปรียบเทียบ ร้อยละ A และ B	เพิ่มขึ้น 3		เพิ่มขึ้น 10		เพิ่มขึ้น 3		เพิ่มขึ้น 10		เท่าเดิม -		เพิ่มขึ้น 14		เพิ่มขึ้น 10		เพิ่มขึ้น 14		เพิ่มขึ้น 27		เพิ่มขึ้น 10		เพิ่มขึ้น 80		เพิ่มขึ้น 36		เพิ่มขึ้น 17		เพิ่มขึ้น 57		เพิ่มขึ้น 21	

*A คือ สร้างระเบียนบรรณานุกรมทั้งหมดภายในการสนทนาเดียว

**B คือ เปิดการสนทนาใหม่ทุกครั้งที่สร้างระเบียนบรรณานุกรมของหนังสือแต่ละชื่อเรื่อง (เลือกใช้เป็นตัวเปรียบเทียบในตารางที่ 2 เนื่องจากได้คะแนนความถูกต้องและโครงสร้างสูงที่สุด)

#C = ป้อนข้อมูลโดยตรง

*** คะแนนเต็ม เท่ากับ 30 คะแนน ส่วนคะแนนเต็มของผลรวมทั้งหมด คือ 420 คะแนน

ตัวเลขในวงเล็บ คือ ร้อยละ

ChatGPT 4-o

ความถูกต้องในภาพรวม คะแนนอยู่ที่ ร้อยละ 54-75 + แต่ละรอบ ผลลัพธ์ไม่เหมือนกัน

– 020 / 040 / 336, 337, 338 / 082 / 100 / 245 / 505 / 520 / 7XX

– เลขผู้แต่ง / 008 / 264 / 6XX

INTERNET REFERENCE SERVICES QUARTERLY
2024, VOL. 28, NO. 3, 287–291
https://doi.org/10.1080/10875301.2024.2361871

Routledge
Taylor & Francis Group

Check for updates

COLUMN

Enriching Bibliographic Records Using AI – A Pilot by Ex Libris

Elizabeth York[a], David Hanegbi[b] and Tamar Ganor[b]

[a]Rutgers University Libraries; Rutgers, The State University of New Jersey; [b]Content Operations, Ex Libris, Part of Clarivate, Jerusalem, Israel

ABSTRACT

This column introduces how Ex Libris uses ChatGPT to enrich ebooks' Bibliographic records with the focus on these MARC fields: 041 (Language), 520 (Summary), and 650 (Subject).

KEYWORDS

Automated cataloging; generative AI; MARC records

https://doi.org/10.1080/10875301.2024.2361871

การจัดทำสาระสังเขปหรือเรื่องย่อ

- หนังสือวิชาการได้ผลลัพธ์ที่ดีกว่า
- ให้ AI อ่าน 15 หน้าแรก → ได้ผลดีกับหนังสือวิชาการ
- ให้ AI อ่านทั้งเล่ม → ได้ผลดีกับนิยาย
- บางครั้ง AI เขียนออกในแนวโฆษณาหนังสือ เช่น

หัวเรื่อง LC

- หนังสือวิชาการได้ผลลัพธ์ที่ดีกว่า
- หัวเรื่องที่ให้บางครั้งไม่ตรงกับในหัวเรื่อง LC
  หรือให้ไม่ตรงกับเนื้อหา เช่น ผิดความหมาย ให้หัวเรื่องแคบไปหรือกว้างไป ฯลฯ

Ex Libris จะพัฒนาต่อไปเรื่อย ๆ และอาจเป็นฟังก์ชันในระบบห้องสมุดฯ

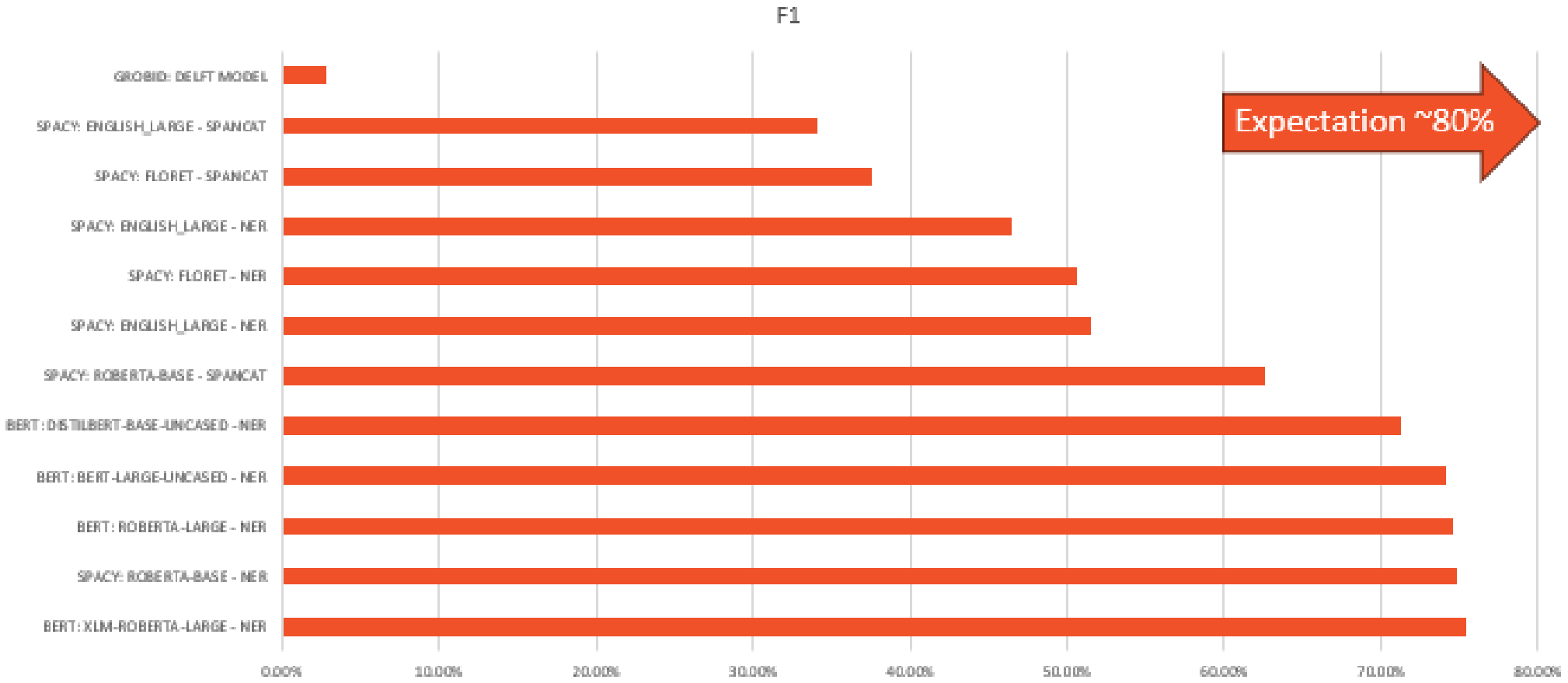
OCLC RLP Webinar
March 12, 2024

Exploring Computational Description: LC Labs Planning Framework in Action

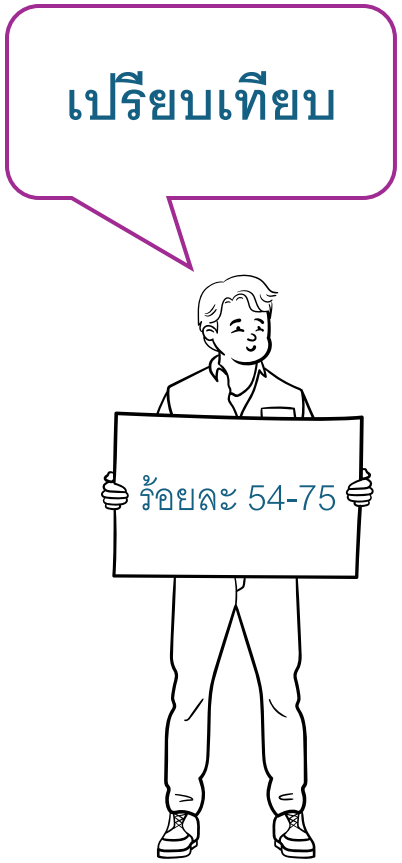
https://www.oclc.org/research/events/2024/ai-planning-framework-in-action.html

Caroline Saccucci & Abigail Potter
Library of Congress
Washington, D.C.

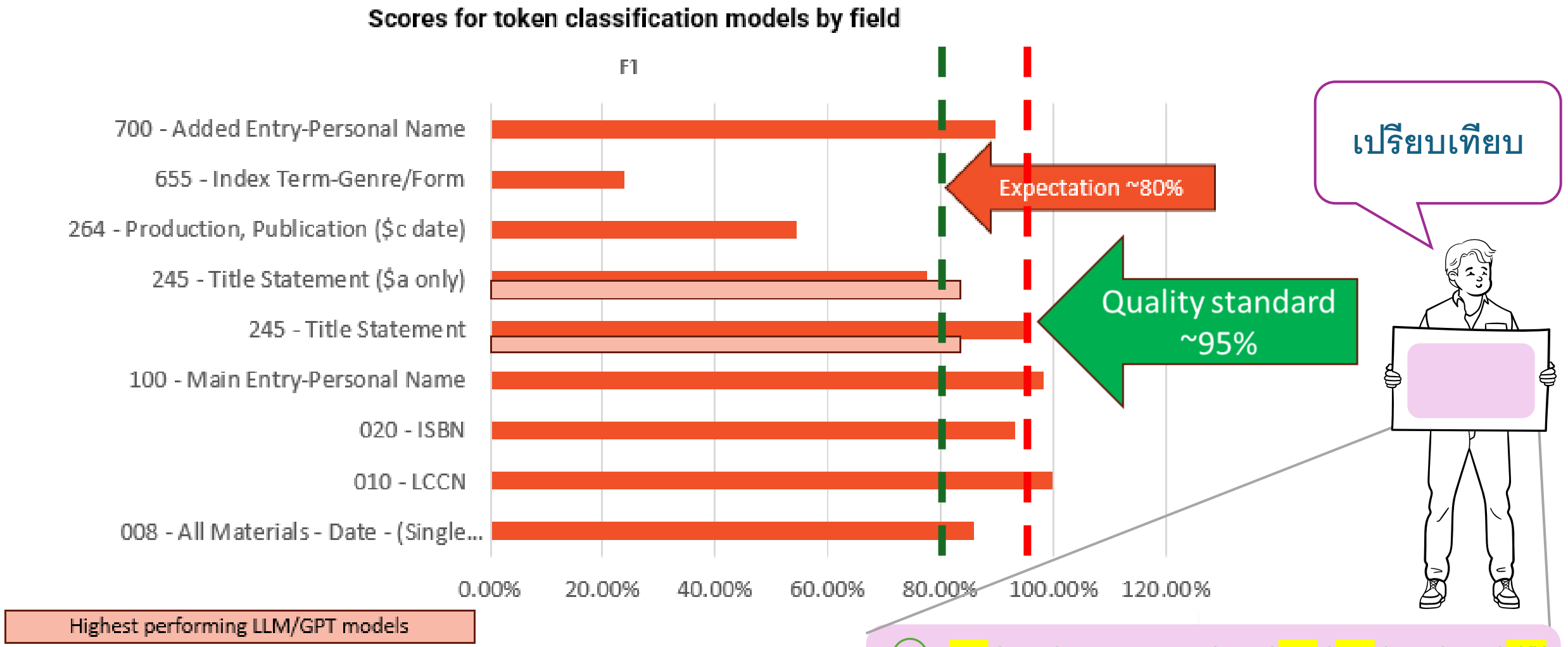
Results: Token Classification for All Fields



F1 score for each of the models, ranked from highest (best), to lowest (worst)



Results: Token Classification for One Field



JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12740

JLIS.it

Annif and Finto AI: Developing and Implementing Automated Subject Indexing

Osma Suominen[(a)], Juho Inkinen[(b)], Mona Lehtinen[(c)]

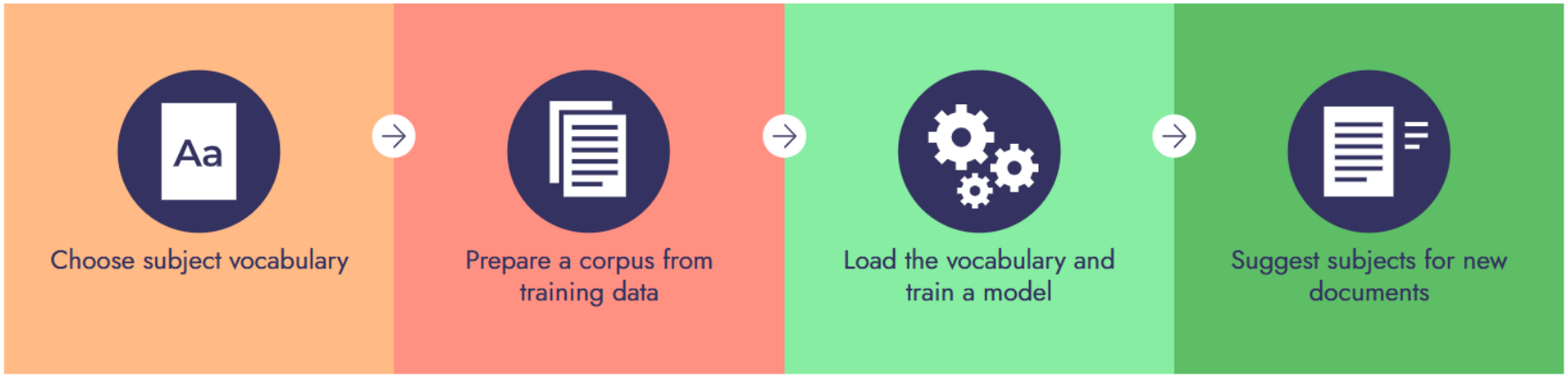
https://www.jlis.it/index.php/jlis/article/view/437

annif

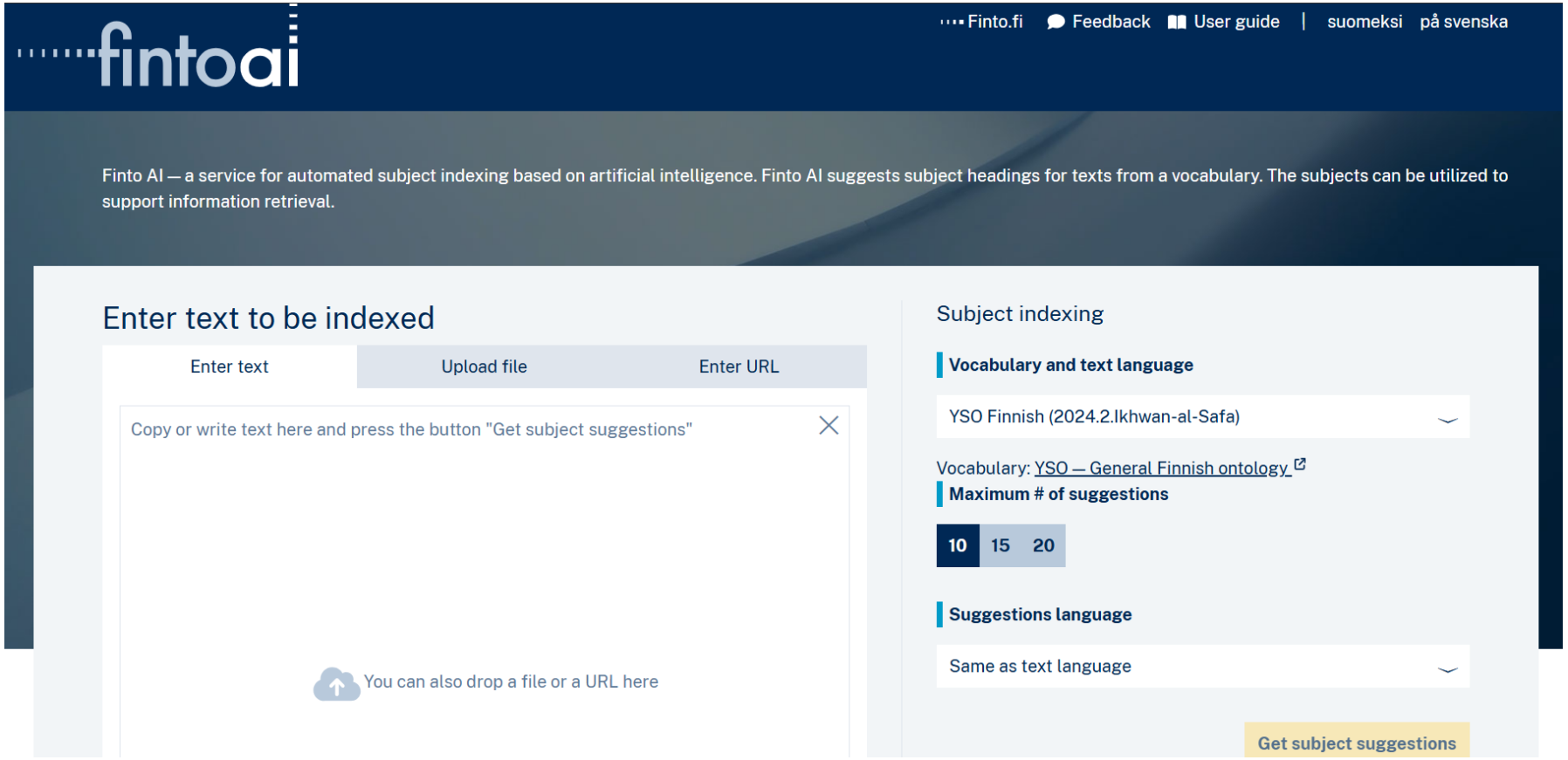
Tool for automated subject indexing and classification

Choose a controlled subject vocabulary and train Annif on already indexed documents – it can then suggest subjects for new documents!

HOW TO USE ANNIF



https://annif.org/



https://ai.finto.fi/

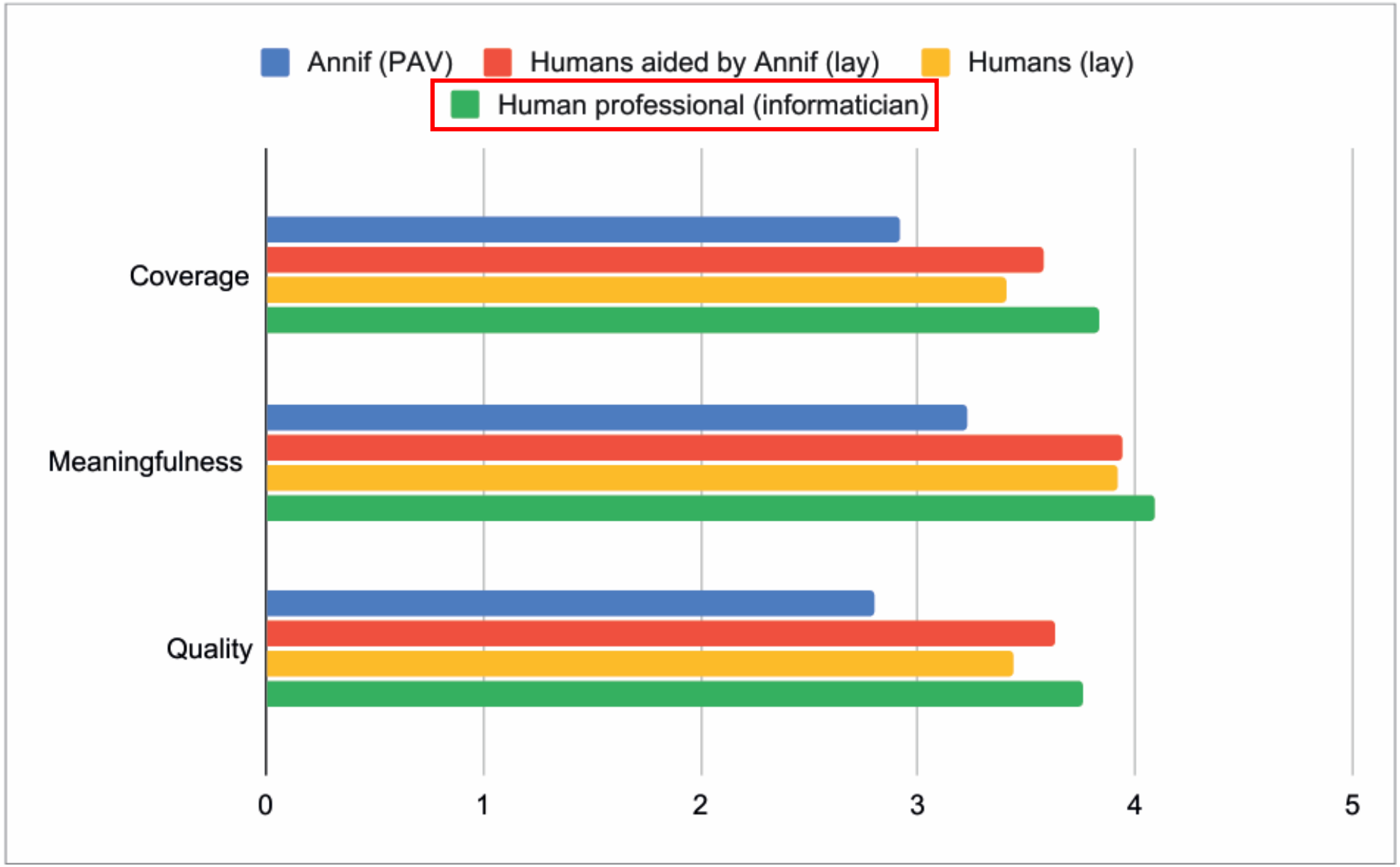


Fig. 3. Quality evaluation of intellectually given and Annif-produced subject indices. Data reproduced from Lehtinen, Inkinen, and Suominen 2019

What did I learn?

1. There are two types of classification in the ML world related to cataloging: **text classification** and **token classification**.
2. One of the models, Annif, had some success with **text classification**: predicting subjects.
3. Some of the models were very successful at **token classification**: predicting authors, titles, and identifiers, such as ISBN and LCCN.
4. ML requires **lots of training data** to improve results, but…
    1. ½ of the training data contained similar patterns of LCSH
    2. ½ of the training data contained unique LCSH
5. Catalogers **reacted more positively to the results** than expected.
6. The **cataloger-assisted prototypes** were really **cool** and have potential.
7. Catalogers are **interested** in ML and seem **less afraid** of it than expected.
8. LLM (ChatGPT) **shows promise** but needs more experimentation.