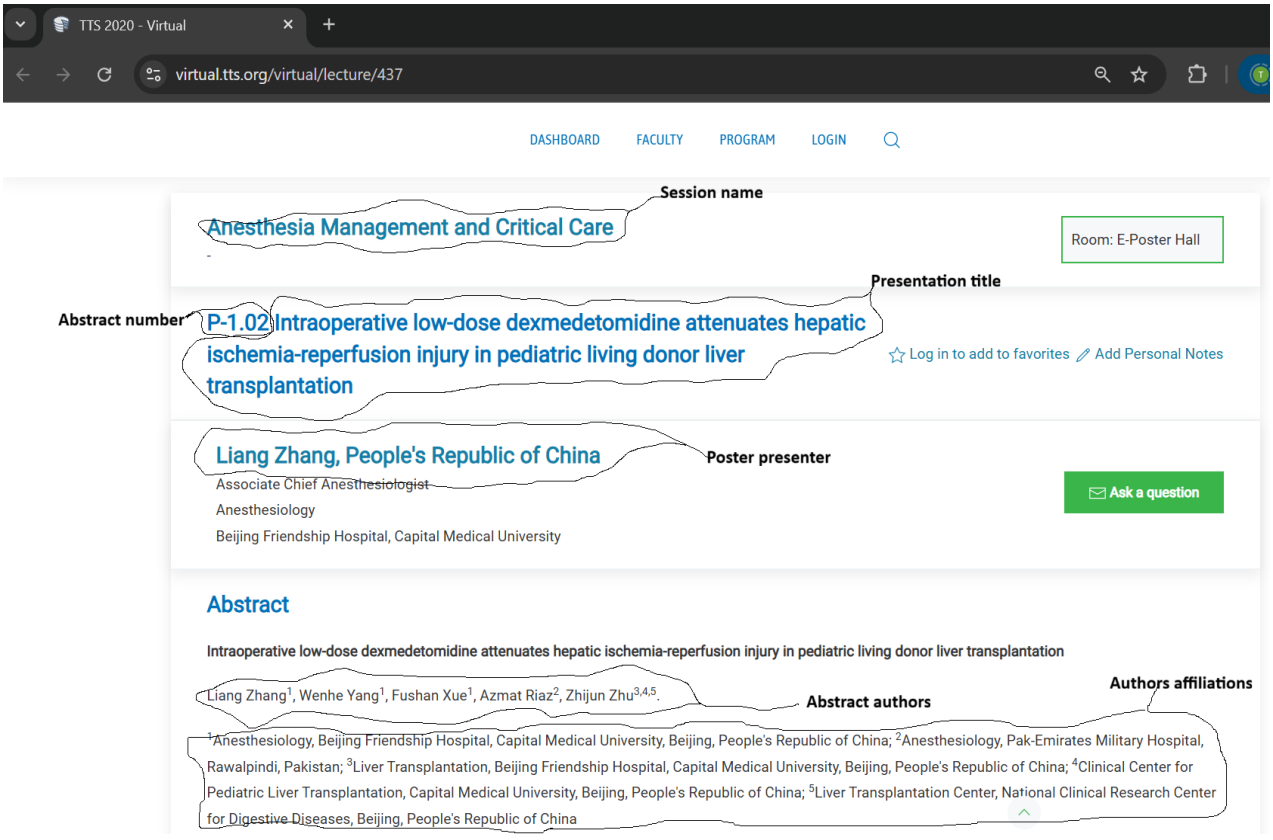
Advanced Content Scraper/Data Engineer at Monocl – Case Assignment

Overall Objective: To transform unstructured information about participants at scientific conferences into structured data. This data will then be used in Monocl SaaS platform (https://monocl.com) . Monocl's clients are pharmaceutical industry players. They use the platform to identify which scientists they should be working with. The frequency at which researchers take part in scientific conferences is often a key factor which speaks of how influential the person might be in the field.

Case Assignment: You need to gather the data from a virtual scientific conference called "28th International Congress of The Transplantation Society" (https://virtual.tts.org/). What we will be using for this assignment however is only the E-Posters (https://virtual.tts.org/virtual/programme/eposters)

In the E-Posters link, you can find a list of links to several poster sessions. You should follow each link to find, gather and parse Presentation Abstracts (see the screenshots below).



Conclusion: Intraoperative low-dose DEX infusion had a potentially protective effect liver graft function in pediatric LDLT without noticeably side effects.

Abstract text

References:

[1] Friedman BH, Wolf JH, Wang L, Putt ME, Shaked A, Christie JD, Hancock WW, Olthoff KM. Serum cytokine profiles associated with early allograft dysfunction in patients undergoing liver transplantation. *Liver Transpl*. 2012; 18: 166-176.

[2] Nickkholgh A, Maluf D. Emerging graft protective strategies in clinical liver transplantation. *Expert Rev Gastroenterol Hepatol*. 2017; 11: 623-631.

[3] Zhai Y, Petrowsky H, Hong JC, Busuttil RW, Kupiec-Weglinski JW. Ischaemia-reperfusion injury in liver transplantation--from bench to bedside. *Nat Rev Gastroenterol Hepatol*. 2013; 11: 79-89.

[4] Sottas CE, Anderson BJ. Dexmedetomidine: the new all-in-one drug in paediatric anaesthesia? *Curr Opin Anaesthesiol*. 2017; 30: 441-451.

[5] Fayed NA, Sayed EI, Saleh SM, Ehsan NA, Elfert AY. Effect of dexmedetomidine on hepatic ischemia-reperfusion injury in the setting of adult living donor liver transplantation. *Clin Transplant*. 2016; 30: 470-482.

[6] Zhang L, Tian M, Xue F, Zhu Z. Diagnosis, Incidence, Predictors and Management of Postreperfusion Syndrome in Pediatric Deceased Donor Liver Transplantation: A Single-Center Study. *Ann Transplant*. 2018; 23: 334-344.

[7] Rosen HR, Martin P, Goss J, Donovan J, Melinek J, Rudich S, Imagawa DK, Kinkhabwala M, Seu P, Busuttil RW, Shackleton CR. Significance of early aminotransferase elevation after liver transplantation. *Transplantation*. 1998; 65: 68-72.

[8] Olthoff KM, Kulik L, Samstein B, Kaminski M, Abecassis M, Emond J, Shaked A, Christie JD. Validation of a current definition of early allograft dysfunction in liver transplant recipients and analysis of risk factors. *Liver Transpl*. 2010; 16: 943-949.

[9] Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO clinical practice guidelines for acute kidney injury. *Kidney Int Suppl*. 2012; 2(Suppl 1): 1-138.

[10] Tokodai K, Lannsjö C, Kjaernet F, Romano A, Januszkiewicz A, Ericzon BG, Nowak G. Association of post-reperfusion syndrome and ischemia-reperfusion injury with acute kidney injury after liver transplantation. *Acta Anaesthesiol Scand*. 2020; Feb 5 [Epub ahead of print].

[11] Guan L, Liu H, Fu P, Li Z, Li P, Xie L, Xin M, Wang Z, Li W. The Protective Effects of Trypsin Inhibitor on Hepatic Ischemia-Reperfusion Injury and Liver Graft Survival. *Oxid Med Cell Longev*. 2016; 2016: 1429835.

DASHBOARD FACULTY PROGRAM LOGIN

Table 1. Baseline Characteristics of Study Patients Before Propensity Score Matching

Variables	Total (n = 115)	Non-DEX (n = 53)	DEX (n = 62)	*P* value
Age (y)	1.3 (0.6-3.8)	1.3 (0.7-3.9)	1.3 (0.6-3.6)	0.649
Male gender (%)	59 (51.3)	28 (52.8)	31 (50.0)	0.762
Height (cm)	75 (67-95)	75 (69-95)	77 (67-96)	0.888
Weight (kg)	10.0 (7.5-15.0)	10.0 (8.0-14.8)	9.6 (7.0-15.0)	0.721
Child score	8 (5-10)	8 (6-10)	7 (5-9)	0.149
PELD score	10 (-7-18)	10 (-7-19)	6 (-8-18)	0.258
Indication for LT (%)				
Biliary atresia	74 (64.3)	35 (66.0)	39 (62.9)	0.726
UCDs[a]	16 (13.9)	7 (13.2)	9 (14.5)	0.840

Task guidelines:

On the individual Abstract page, you should **extract** the following:

- session name,
- abstract number,
- presentation title,
- Person's name,
- Person's conference role ("Poster presenter" or "Abstract author")
- person's affiliations
- Abstract text
- Abstract URL

If Abstract text contains any tables – skip those.

Make sure to include References list (if available) in the abstract text.

To match a person to their affiliation you should use a superscript number which you can see following each name. If the person has more than one affiliation – aggregate the organization names in a list.

Data is already filled in for the presentation P-1.02 and your task is to continue for the next couple of presentations. Make sure to append the data you scraped to the existing file. Use Pandas library for those purposes.

Submit **two** files as a result of this task:

File_1:

- append the data you scraped to the existing file.
- ensure the file has at least 2 poster sessions.
- Ensure you write one person's name per row

File_2:

- Filter the data so the output file contains only the people who have **Abstract author** as their role.
- Split the name column into fisrtName, middleName, lastName and Title (e.g. MD, PhD etc) fields. You can assume that the names are written in this order – first name, middle name, and last name.

Final remarks:

There is more than one way to solve this task. You should be able to justify and explain the approach you chose. As part of this assignment, we will then sit down with you, discuss your approach to this problem, and look at the solution you used for the problem at hand.

As much as possible, while working on this task we expect you to extract the data through coding/scraping rather than manual steps (e.g. we encourage you to use Scrapy for scraping, and Pandas for working with data).