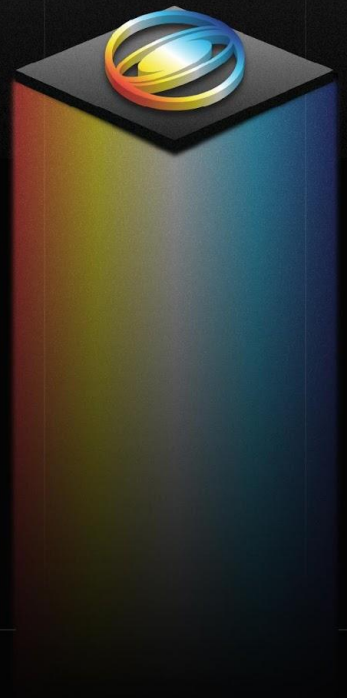
SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

THE PREMIER CONFERENCE & EXHIBITION ON COMPUTER GRAPHICS & INTERACTIVE TECHNIQUES

(Tiny) Problems with Web Scraping the ACM Digital Library



SIGGRAPH 2024 Village: History: The ACM SIGGRAPH History Archive—Moving Forward by Looking Backwards

Manuel Alducin

Information Technology Professional

Global Affairs Canada

At the Embassy of Canada to Mexico

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Web Scraping

Process

Download

Save some time

- Download the Publication pages
- Use filename for some configuration

Extract Info

Two Steps

- Get basic info from Publication page
- Get Abstract and (rebuild) References from detail page

Write Excel

Write File

- Except no buffering, write until the end!

Setup

- Ubuntu 22
- Python 3.10 (tried 3.11 for TOML via deadsnakes)
- Main libraries
  - BeautifulSoup: parse the HTML
  - XlsxWriter: directly write Excel file
  - requests: get detail page
- Script about 150 lines of code
- No logging (only print to screen)
- Configuration inside the script
- Script more or less worked within 2 weeks
- No Git! (what was I thinking!)

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

The Main Problem

Time

The Main Problem: Time

Late Involvement

- Learned of History Project at Vancouver 2022
- Got in touch around December 2022

Web Scraping

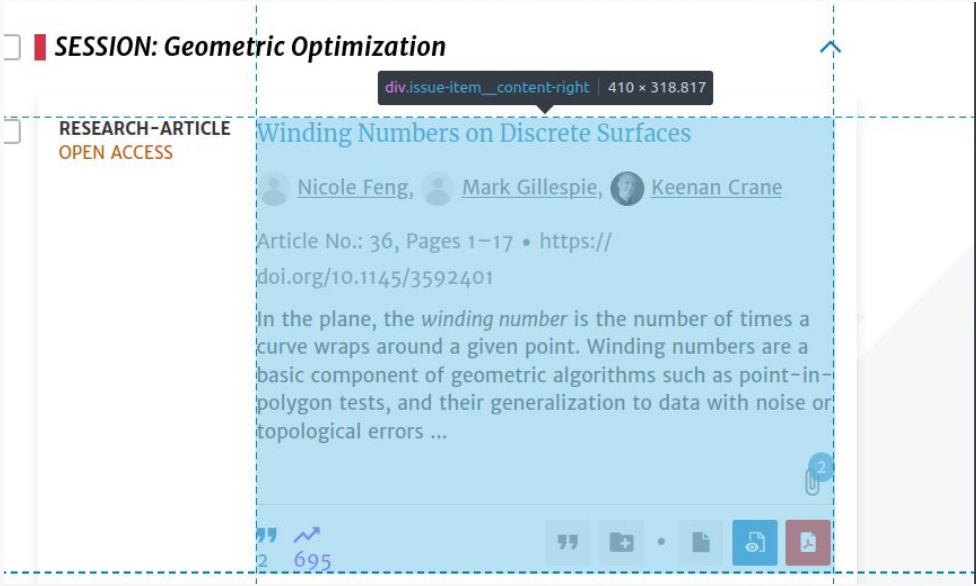
- Can still require manual cleanup
- May require lots of tweaking
- Hadn’t done scraping in a while

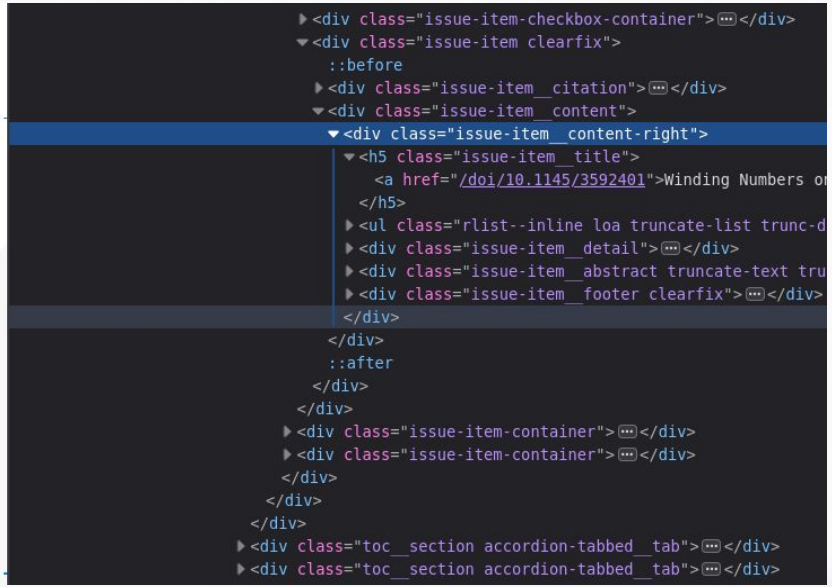
Analyse the Problem

- Little time to really analyze the DL in-depth
- But the DL seems pretty clean, right?

Scraping the DL

SIGGRAPH 2023 TOG





The Final Output

SIGGRAPH 2023 Papers

learning_title	learning_conference	rel_learning_type	learning_abstract	learning_link_to_dl	learning_references	Presenters
	rel:conference_shor	tax:learning_type				
Winding Numbers on Discrete Surfaces	SIGGRAPH 2023	Technical Papers	In the plane, the winding number is the number of times a curve	https://doi.org/10.1145/3592401	65. Qingnan Zhou, Eitan Grinspun, Denis Zorin, and Alec Jacobson. 2016. Me	Keenan Crane
Flexible Isosurface Extraction for Gradient-Based Mesh Optimizat	SIGGRAPH 2023	Technical Papers	This work considers gradient-based mesh optimization, where we	https://doi.org/10.1145/3592430	74. Chenyang Zhu, Kai Xu, Siddhartha Chaudhuri, Renjiao Yi, and Hao Zhan	Jun Gao
Topology driven approximation to rational surface-surface interse	SIGGRAPH 2023	Technical Papers	Computing the intersection between two parametric surfaces (SS	https://doi.org/10.1145/3592452	63. Y Yamaguchi, R Kamiyama, and F Kimura. 2001. Surface-Surface Interse	Ming Li
A Practical Wave Optics Reflection Model for Hair and Fur	SIGGRAPH 2023	Technical Papers	Traditional fiber scattering models, based on ray optics, are miss	https://doi.org/10.1145/3592446	57. Arno Zinke, Martin Rump, Tomás Lay, Andreas Weber, Anton Andriyenk	Steve Marschner
Anatomically Detailed Simulation of Human Torso	SIGGRAPH 2023	Technical Papers	Many existing digital human models approximate the human ske	https://doi.org/10.1145/3592425	59. Jānis Šavlovskis and Kristaps Raits. [n. d.]. Biomechanics of the Spine. h	C. Karen Liu
HACK: Learning a Parametric Head and Neck Model for High-fide	SIGGRAPH 2023	Technical Papers	Significant advancements have been made in developing parame	https://doi.org/10.1145/3592093	115. Yiyu Zhuang, Hao Zhu, Xusen Sun, and Xun Cao. 2022. Mofanerf: Morp	Jingyi Yu
GestureDiffuCLIP: Gesture Diffusion Model with CLIP Latents	SIGGRAPH 2023	Technical Papers	The automatic generation of stylized co-speech gestures has rece	https://doi.org/10.1145/3592097	88. Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On	Libin Liu
BodyFormer: Semantics-guided 3D Body Gesture Synthesis with	SIGGRAPH 2023	Technical Papers	Automatic gesture synthesis from speech is a topic that has attra	https://doi.org/10.1145/3592456	51. Yukun Zhu, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel	Taku Komura
Listen, Denoise, Action! Audio-Driven Motion Synthesis with Diffu	SIGGRAPH 2023	Technical Papers	Diffusion models have experienced a surge of interest as highly e	https://doi.org/10.1145/3592458	133. Vikram Voleti, Alexia Jolicoeur-Martineau, and Christopher Pal. 2022. M	Gustav Eje Henter
Contact Edit: Artist Tools for Intuitive Modeling of Hand-Object In	SIGGRAPH 2023	Technical Papers	Posing high-contact interactions is challenging and time-consumi	https://doi.org/10.1145/3592117	43. W. Zhao, J. Zhang, J. Min, and J. Chai. 2013. Robust Realtime Physics-Ba	Nancy Pollard
Eventfulness for Interactive Video Alignment	SIGGRAPH 2023	Technical Papers	Humans are remarkably sensitive to the alignment of visual ever	https://doi.org/10.1145/3592118	50. Oliver Wang, Christopher Schroers, Henning Zimmer, Markus Gross, and	Abe Davis
FactorMatte: Redefining Video Matting for Re-Composition Tasks	SIGGRAPH 2023	Technical Papers	We propose Factor Matting, an alternative formulation of the vid	https://doi.org/10.1145/3592423	46. Yunke Zhang, Chi Wang, Miaomiao Cui, Peiran Ren, Xuansong Xie, Xian	Abe Davis
Computational Long Exposure Mobile Photography	SIGGRAPH 2023	Technical Papers	Long exposure photography produces stunning imagery, represer	https://doi.org/10.1145/3592124	46. Timo Zinßer, Jochen Schmidt, and Heinrich Niemann. 2005. Point set reg	Yael Pritch
ShapeCoder: Discovering Abstractions for Visual Programs from U	SIGGRAPH 2023	Technical Papers	We introduce ShapeCoder, the first system capable of taking a da	https://doi.org/10.1145/3592416	57. Fenggen Yu, Zhiqin Chen, Manyi Li, Aditya Sanghi, Hooman Shayani, Ali	Daniel Ritchie
The Visual Language of Fabrics	SIGGRAPH 2023	Technical Papers	We introduce text2fabric, a novel dataset that links free-text des	https://doi.org/10.1145/3592391	70. Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Kalyan S	Belen Masia
ArrangementNet: Learning Scene Arrangements for Vectorized In	SIGGRAPH 2023	Technical Papers	We present a novel vectorized indoor modeling approach that co	https://doi.org/10.1145/3592122	85. Chuhang Zou, Ersin Yumer, Jimei Yang, Duygu Ceylan, and Derek Hoiem	Li Yi
Juxtaform: interactive visual summarization for exploratory shap	SIGGRAPH 2023	Technical Papers	We present juxtaform, a novel approach to the interactive summ	https://doi.org/10.1145/3592436	70. Jun-Yan Zhu, Yong Jae Lee, and Alexei A Efros. 2014. Averageexplorer: In	Karan Singh
Patternshop: Editing Point Patterns by Image Manipulation	SIGGRAPH 2023	Technical Papers	Point patterns are characterized by their density and correlation,	https://doi.org/10.1145/3592418	93. Yang Zhou, Zhen Zhu, Xiang Bai, Dani Lischinski, Daniel Cohen-Or, and	Gurprit Singh
VideoDoodles: Hand-Drawn Animations on Videos with Scene-Aw	SIGGRAPH 2023	Technical Papers	We present an interactive system to ease the creation of so-calle	https://doi.org/10.1145/3592413	64. Zhoutong Zhang, Forrester Cole, Richard Tucker, William T Freeman, an	Adrien Bousseau
StripMaker: Perception-driven Learned Vector Sketch Consolidati	SIGGRAPH 2023	Technical Papers	Artist sketches often use multiple overdrawn strokes to depict a s	https://doi.org/10.1145/3592130	62. Jerry Yin, Chenxi Liu, Rebecca Lin, Nicholas Vining, Helge Rhodin, and A	Alla Sheffer
Semi-supervised reference-based sketch extraction using a contr	SIGGRAPH 2023	Technical Papers	Sketches reflect the drawing style of individual artists; therefor	https://doi.org/10.1145/3592392	79. Changqing Zou, Haoran Mo, Chengying Gao, Ruofei Du, and Hongbo Fu	Junyong Noh
Split-Lohmann Multifocal Displays	SIGGRAPH 2023	Technical Papers	This work provides the design of a multifocal display that can cre	https://doi.org/10.1145/3592110	42. Tao Zhan, Jianghao Xiong, Junyu Zou, and Shin-Tson Wu. 2020. Multifoca	Aswin C. Sankaranarayanan
Étendue Expansion in Holographic Near Eye Displays through Spa	SIGGRAPH 2023	Technical Papers	In this paper, we present a novel method the étendue expansion	https://doi.org/10.1145/3592441	33. Wei Yuan, Li-Hua Li, Wing-Bun Lee, and Chang-Yuen Chan. 2018. Fabric	Yoonchan Jeong
Rhizomorph: The Coordinated Function of Shoots and Roots	SIGGRAPH 2023	Technical Papers	Computer graphics has dedicated a considerable amount of effor	https://doi.org/10.1145/3592145	78. Y. Zhao and J. Barbič. 2013. Interactive Authoring of Simulation-ready Pl	Wojtek Pałubicki

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Challenges

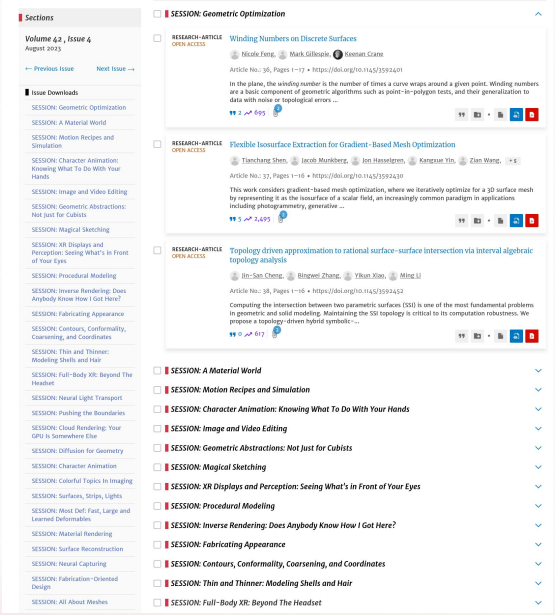
Discrepancies in the DL

Challenges: Discrepancies in the DL

- DL is rate limited: find out how low you can go
- Updates to the DL System
- CSS specificity
- Differences between TOG and older Proceedings
- Older Proceedings might include info from several types of Sessions: Panels, Sketches, etc.

Challenges: Discrepancies in the DL

SIGGRAPH 2023 TOG



SIGGRAPH 1995 Proceedings

ARTICLE
FREE

Computer graphics achievement award

Kurt Akeley

Page 10 • https://doi.org/10.1145/218380.218386

0 417

ARTICLE
FREE

Steven A. Coons award for outstanding creative contributions to computer graphics

Jose Luis Encarnação

Pages 11–12 • https://doi.org/10.1145/218380.218388

0 356

ARTICLE
FREE

Geometry compression

Michael Deering

Pages 13–20 • https://doi.org/10.1145/218380.218391

431 2,996

ARTICLE
FREE

Polygon-assisted JPEG and MPEG compression of synthetic images

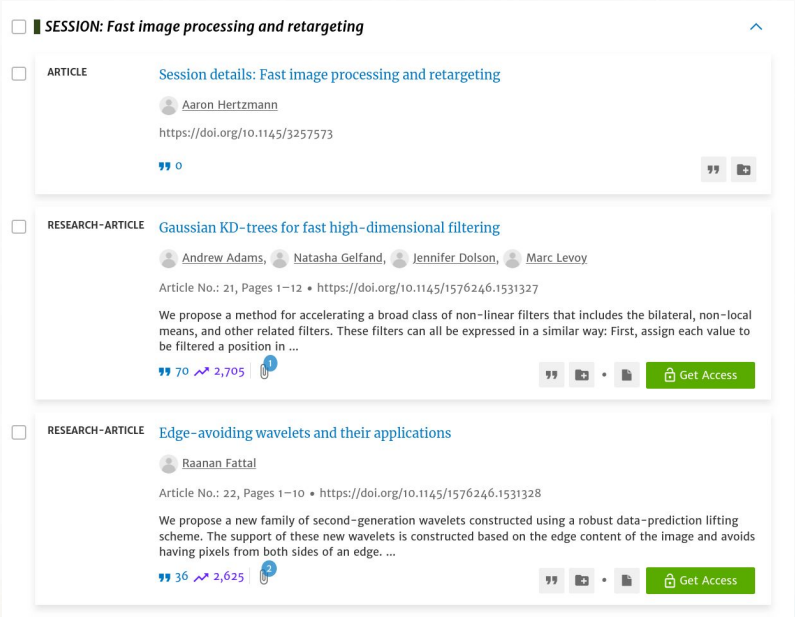
Marc Levoy

Pages 21–28 • https://doi.org/10.1145/218380.218392

58 1,198

Challenges: Discrepancies in the DL

SIGGRAPH 2009 Papers



SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Challenges

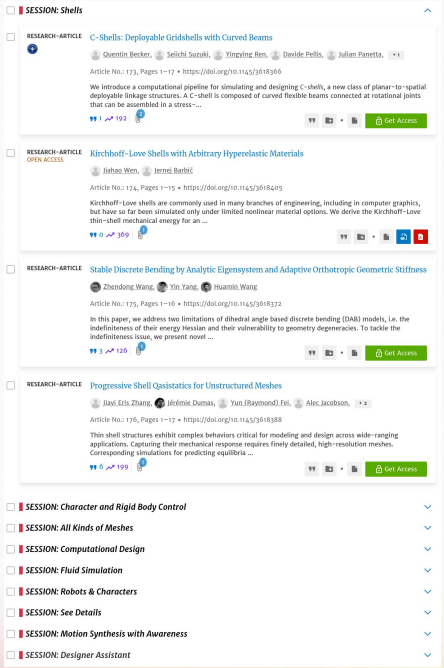
Dynamic Content

Challenges: Dynamic Content

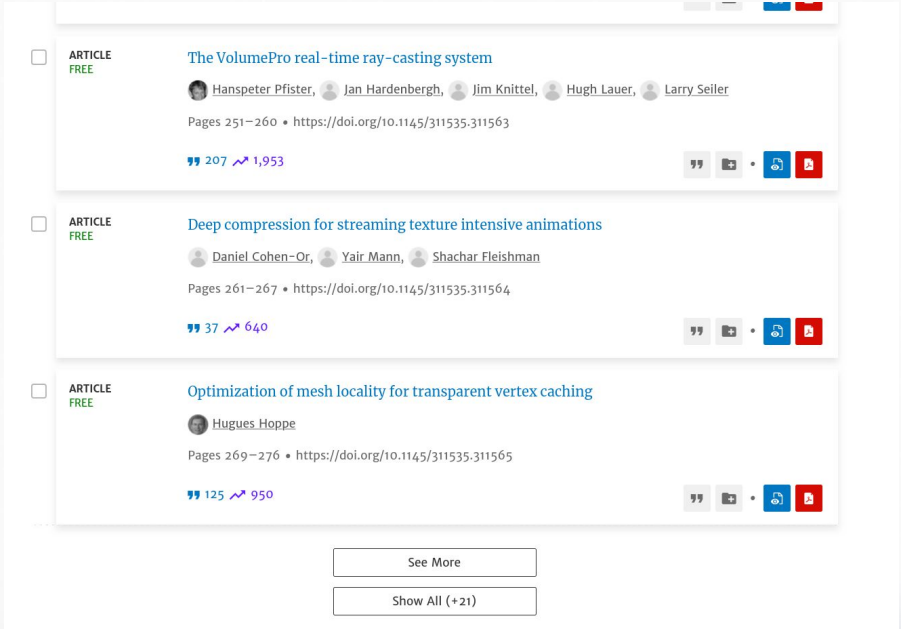
- Used for large lists of Articles and for large number of authors
- Also different between older and newer content
- Two options: try to automate or take care manually
- Automate: use Selenium?
- Manually: just click away

Challenges: Dynamic Content

SIGGRAPH Asia 2023



SIGGRAPH 1999 Proceedings



Challenges: Dynamic Content

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

SIGGRAPH Asia 2023

RESEARCH-ARTICLE
OPEN ACCESS

SLANG.D: Fast, Modular and Differentiable Shader Programming

Sai Praveen Bangaru, Lifan Wu, Tzu-Mao Li, Jacob Munkberg, Gilbert Bernstein, + 4

Article No.: 264, Pages 1–28 • https://doi.org/10.1145/3618353

We introduce SLANG.D, an extension to the Slang shading language that incorporates first-class automatic differentiation support. The new shading language allows us to transform a Direct3D-based path tracer to be fully differentiable with minor ...

1 428 1

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Challenge

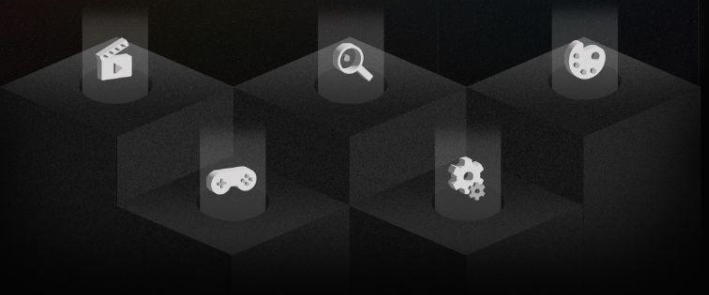
More Time Pressure!

Challenge: More Time Pressure!

- Early 2023: need citations of all Papers for the Seminal Graphics Papers Vol. 2
- How do citations work in the DL?
- It’s dynamic content but easy to get from page
- Re-download all Proceedings and TOG volumes again for up to date
- Re-process all HTML to get citations
- Either send the new Excels or try to merge manually

Bumps

- Borked system (deadsnakes!): reinstall and switch to miniconda (Python 3.12 with TOML)
- Papers are now separated between TOG and Proceedings



SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Thanks!