Search For E-Commerce: (*Not*) Solved (*Yet*)

June 12, 2018

Liangjie Hong
Head of Data Science, Etsy Inc.

SIGIR eCom'18 Workshop

Liangjie Hong

- **Head of Data Science** at **Etsy** since Aug, 2016
  Machine Learning Solutions for **Search**, **Recommendation** and **Computational Advertising**
  Offices in New York City, San Francisco and Toronto

- **Senior Manager of Research** at **Yahoo Research** in Sunnyvale, CA (2013 – 2016)
  Leading science efforts for personalization and search sciences

---

- Multi-modal Deep-learning based Search Solution (KDD 2016)
- Probabilistic Graphical Model based Personalization Recommendation (KDD 2014)
- Ensemble Learning based CTR Prediction Solution (AdKDD 2017/KDD 2017)
- Buzzsaw: A System for High Speed Feature Engineering (SysML 2018)
- Learning Within-Session Budgets from Browsing Trajectories for Item Recommendations (RecSys 2018)

Agenda

- **An Introduction to Etsy**
- **Challenges to Search for E-Commerce**
- **Etsy’s Efforts on Search Ranking**

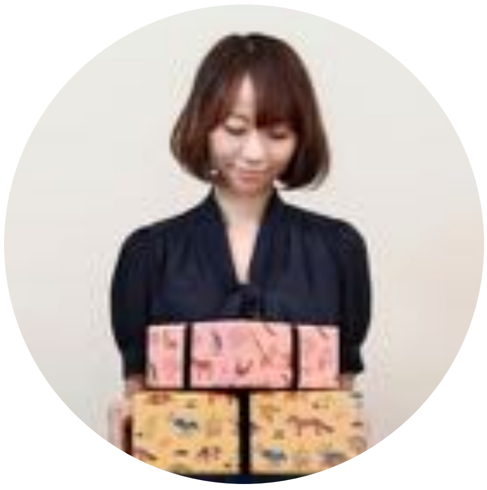
An Introduction to Etsy

OUR MISSION

Reimagine commerce in ways that build a more fulfilling and lasting world

Etsy – A Global Marketplace


Artifact Bags
Omaha, NE
Photo by: Dana Damewood and Jackie Sterba


Clap Clap
Los Angeles, CA
Photo by: Bert Youn and Mimi Kim


redravenstudios
Pittsburgh, PA
Photo by: Janelle Bendycki


Little Hero Capes
Somerset, MA
Photo by: Rich Vintage Photography


Cattails Woodwork
Hermitage, PE, Canada
Photo by: Cattails Woodwork


Room for Emptiness
Berlin, Germany
Photo by: Room for Emptiness


sukrachand
Brooklyn, NY
Photo by: sukrachand


Nicole Porter Design
Saint Paul, MN
Photo by: Nicole Porter Design


noemiah
Montreal, QC, Canada
Photo by: noemiah


Lorgie
Fremantle, WA, Australia
Photo by: Lorgie


Jeremiah Collection
San Francisco, CA
Photo by: Matthew Reamer


Docksmith
Brunswick, ME
Photo by: Docksmith

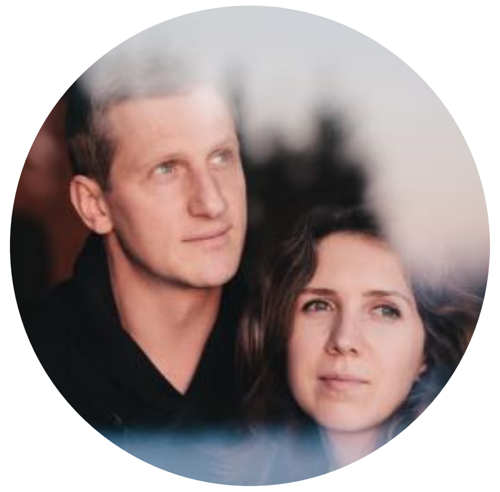

purlBKnit
Brooklyn, NY
Photo by: purlBKnit


Julia Astreou
Nicosia, Cyprus
Photo by: Panagiotis Mina


Moira K. Lime
Omaha, NE
Photo by: Moira K. Lime


Nested Yellow
Portland, OR
Photo by: Jessica Dremov and Nested Yellow


Habitables
Madrid, Spain
Photo by: Habitables


Woodstorming
Kaunas, Lithuania
Photo by: Ilona & Martynas from Instudija


karoArt
Dublin, Ireland
Photo by: Christine Burns


ADIKILAV
Jerusalem, Israel
Photo by: Shlomit Koslowe

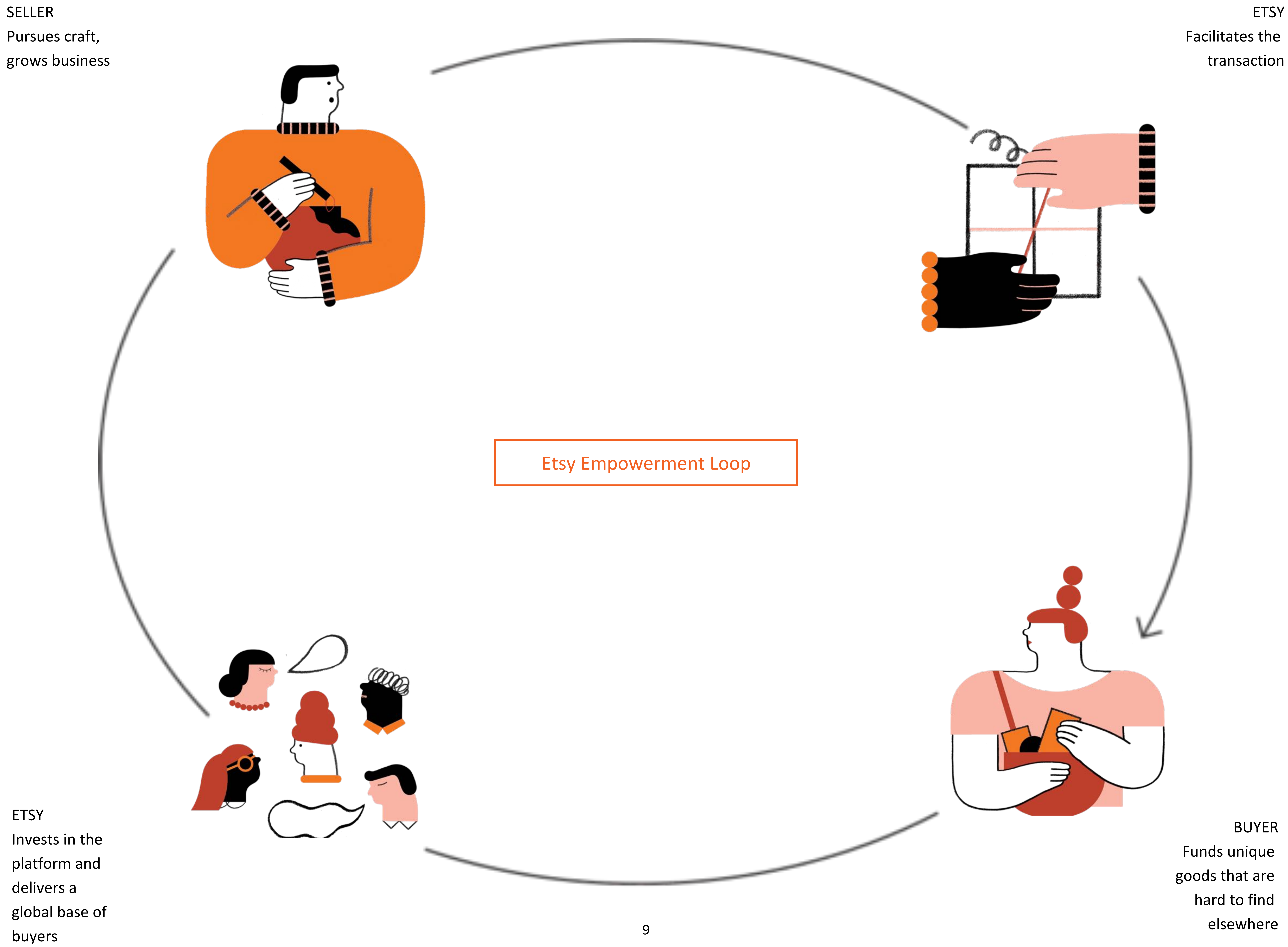

My A La Mode Boutique
Ecuador
Photo by: My A La Mode Boutique

Etsy – A Global Marketplace

By The Numbers

1.9M	31.7M	$2.8B	45+M
active sellers	active buyers	annual GMS	items for sale

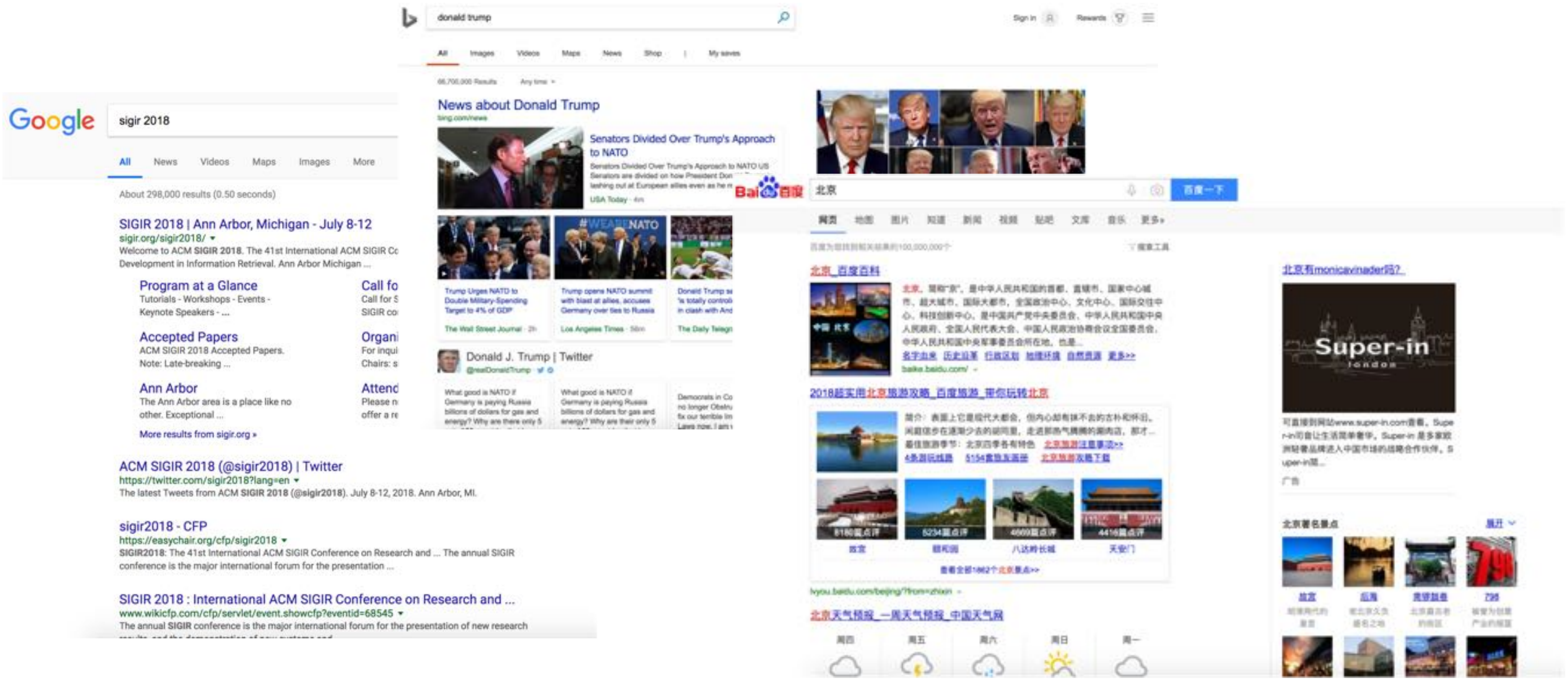

Photo by Kirsty-Lyn Jameson


SELLER
Pursues craft, grows business
ETSY
Facilitates the transaction
Etsy Empowerment Loop
BUYER
Funds unique goods that are hard to find elsewhere
ETSY
Invests in the platform and delivers a global base of buyers

Challenges to Search for E-Commerce

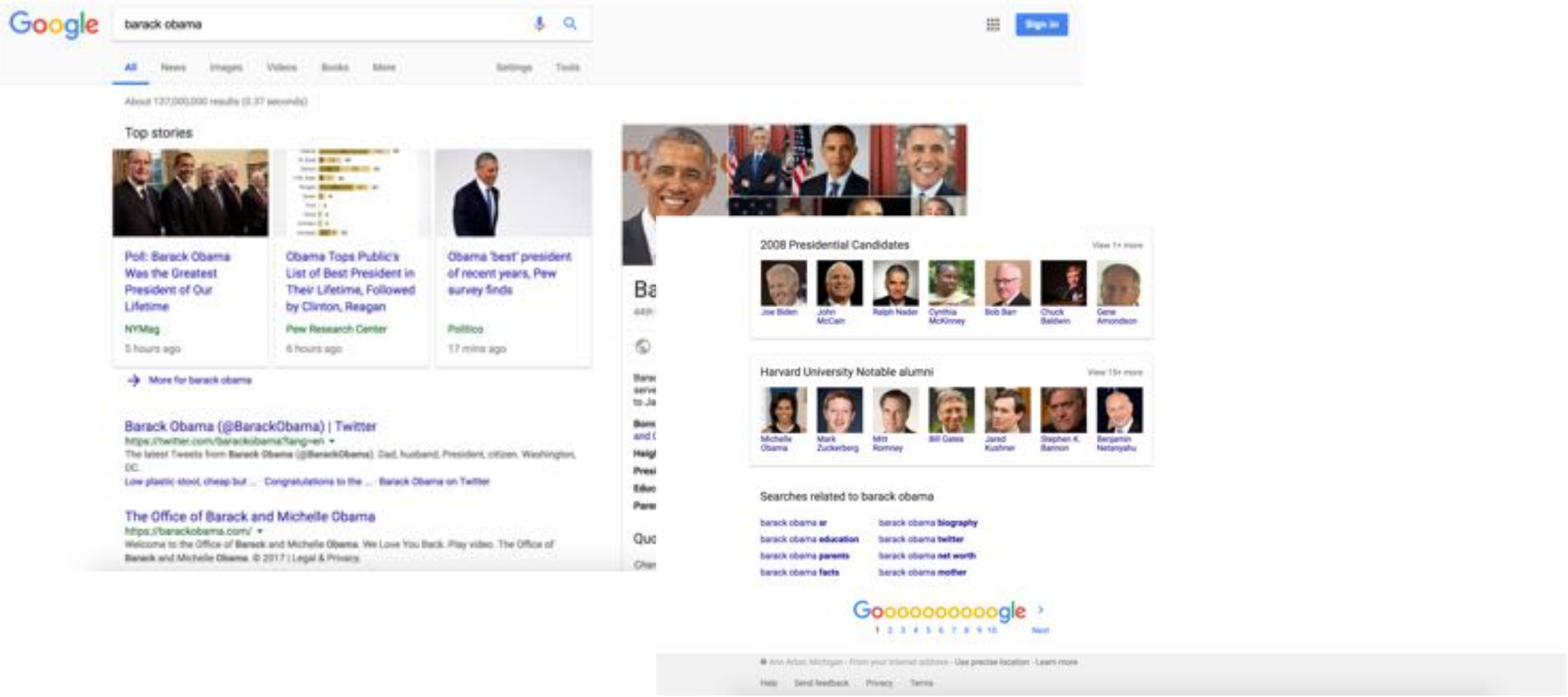
Challenges to Search for E-Commerce

Generic Search



Challenges to Search for E-Commerce

Generic Search



Challenges to Search for E-Commerce

Generic Search

- **Classic Information Retrieval (1950 – 1990)**
  TF-IDF, BM25, Language Models

- **Learning To Rank (2000 – 2010)**
  RankSVM, GBDT, LambdaMART

- **Neural Learning To Rank (2013 – Present)**
  DSSM, DESM, IRGAN

Challenges to Search for E-Commerce

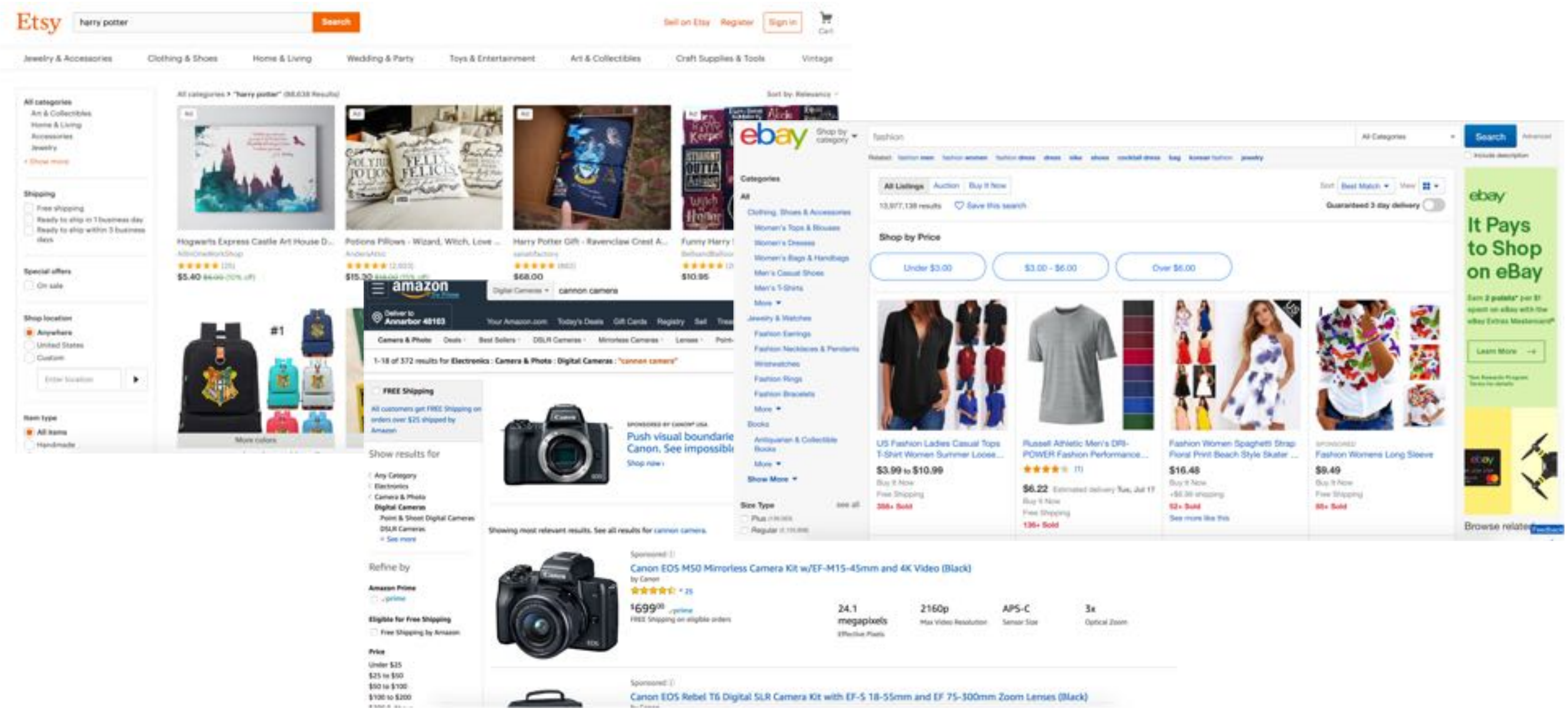
Generic Search

- **Cranfield Paradigm and Test Collections (1950 – Present)**
  TREC (1992 – Present)
  Microsoft Learning To Rank (2007 – 2009)
  Yahoo Learning To Rank Challenge (2011)

- **Understanding Implicit Feedback and Relevance (2000 – Present)**
  Thorsten Joachims's Work
  Eugene Agichtein's Work

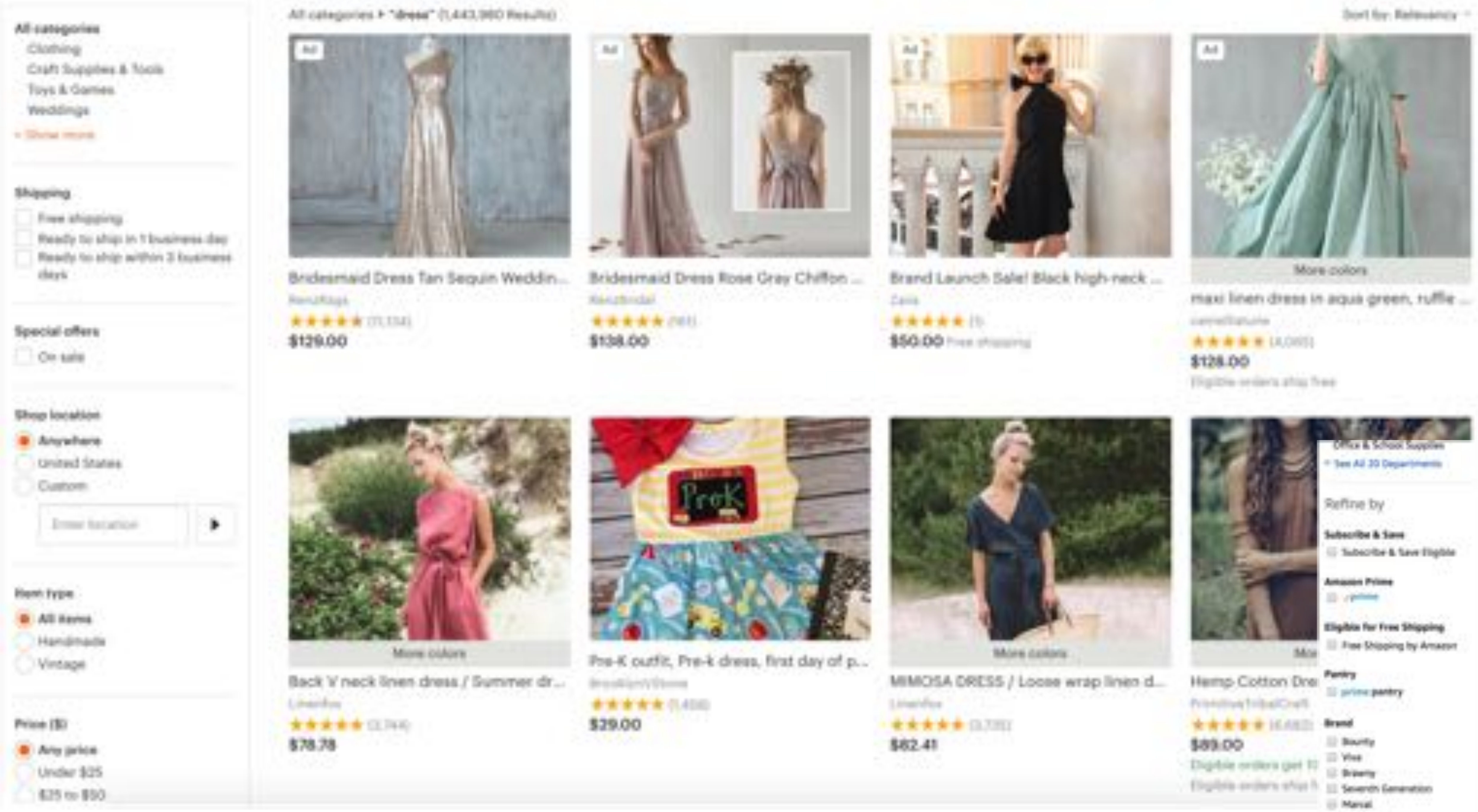
Challenges to Search for E-Commerce

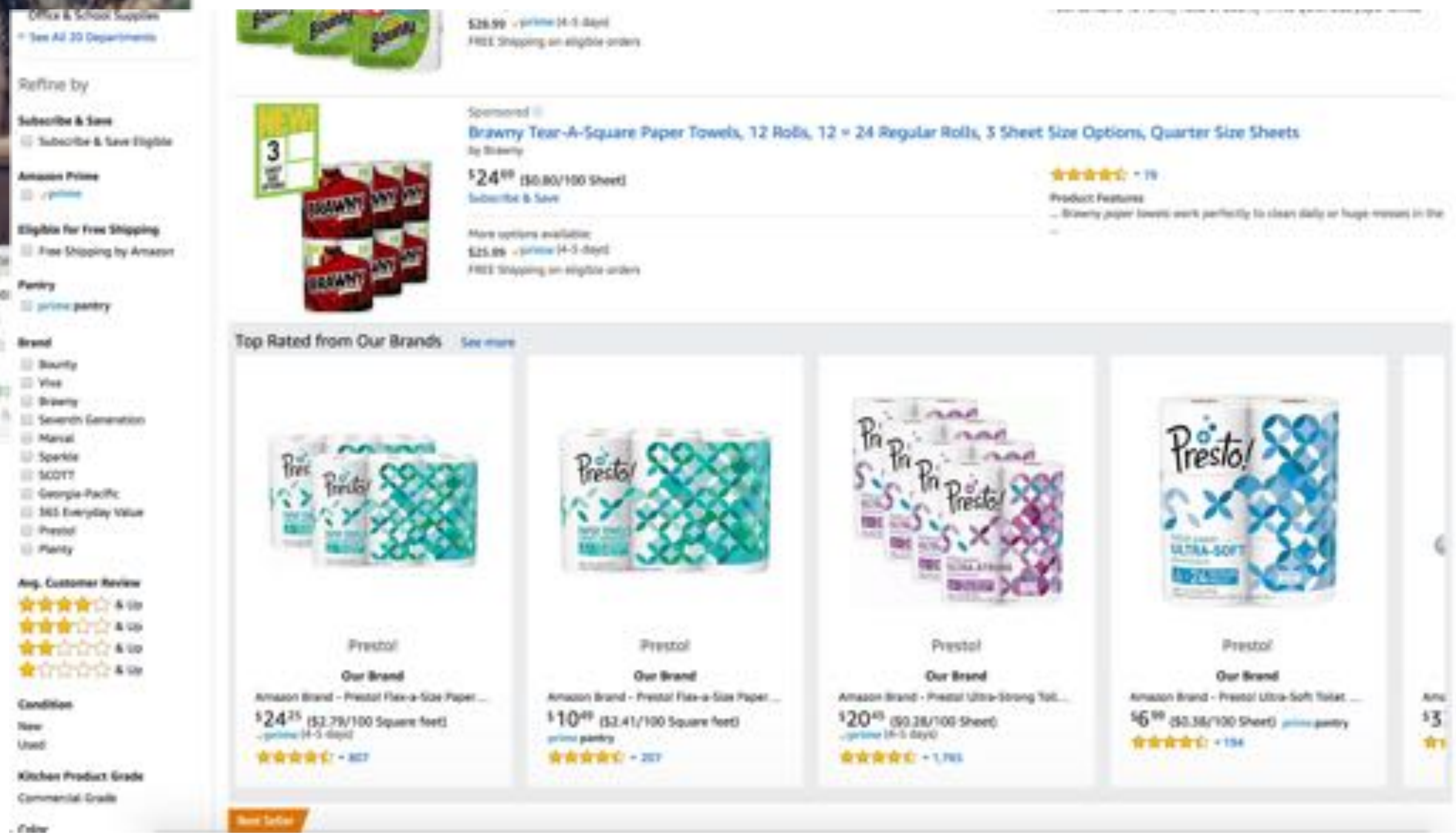
Search for E-Commerce



Challenges to Search for E-Commerce

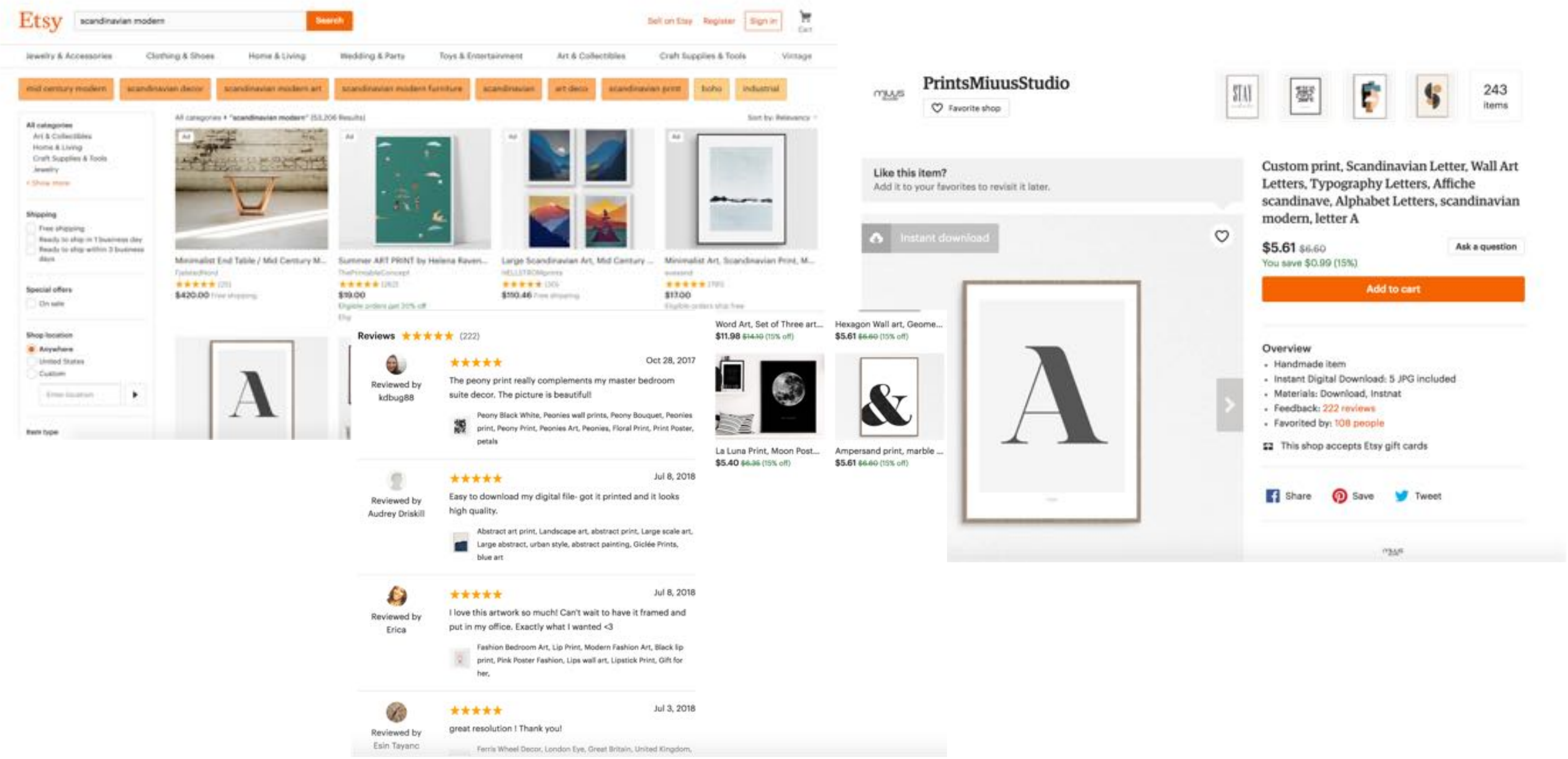
Challenge I - Relevancy





Challenges to Search for E-Commerce

Challenge I - Relevancy

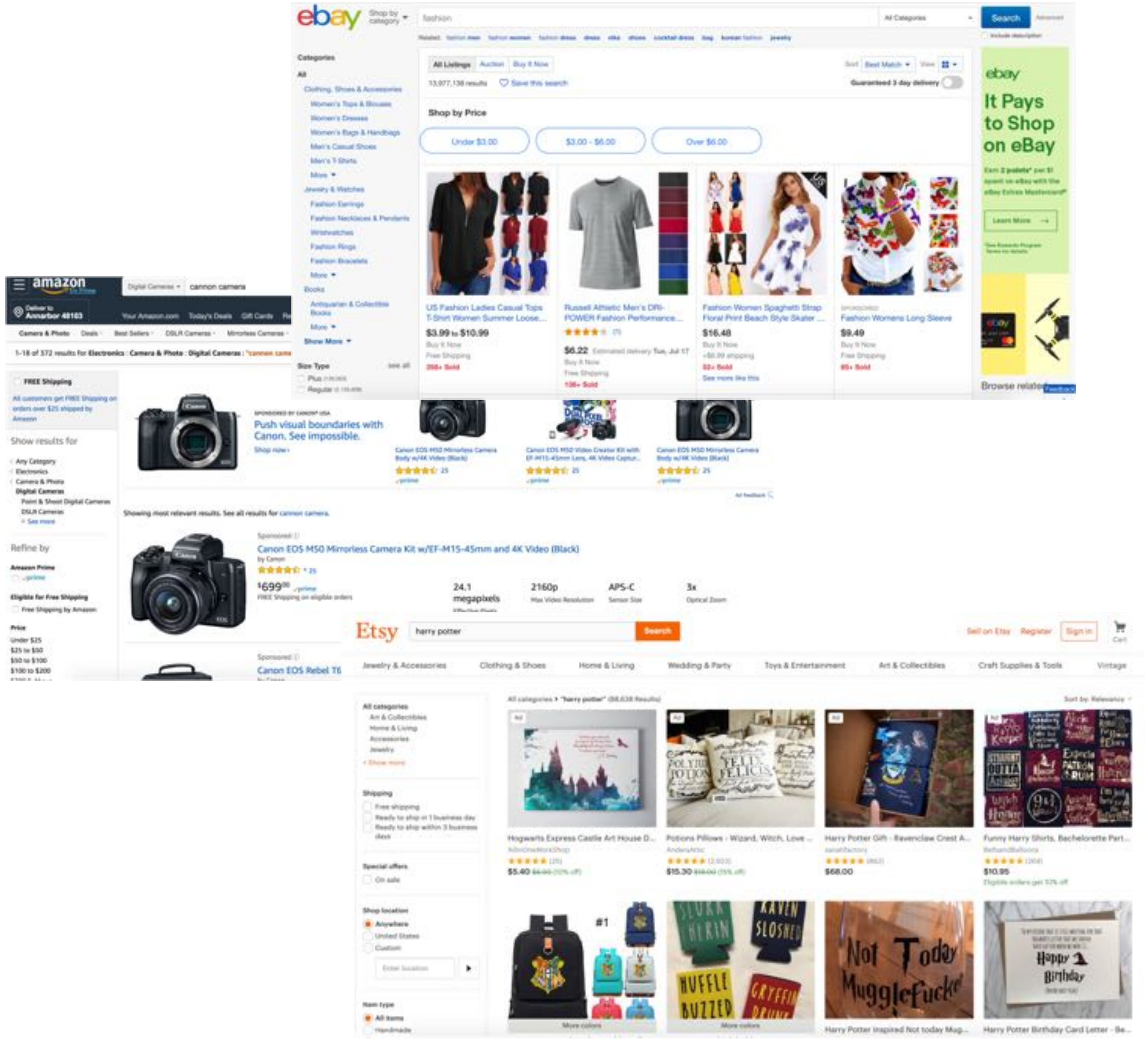


Challenges to Search for E-Commerce

Challenge II – User Satisfaction

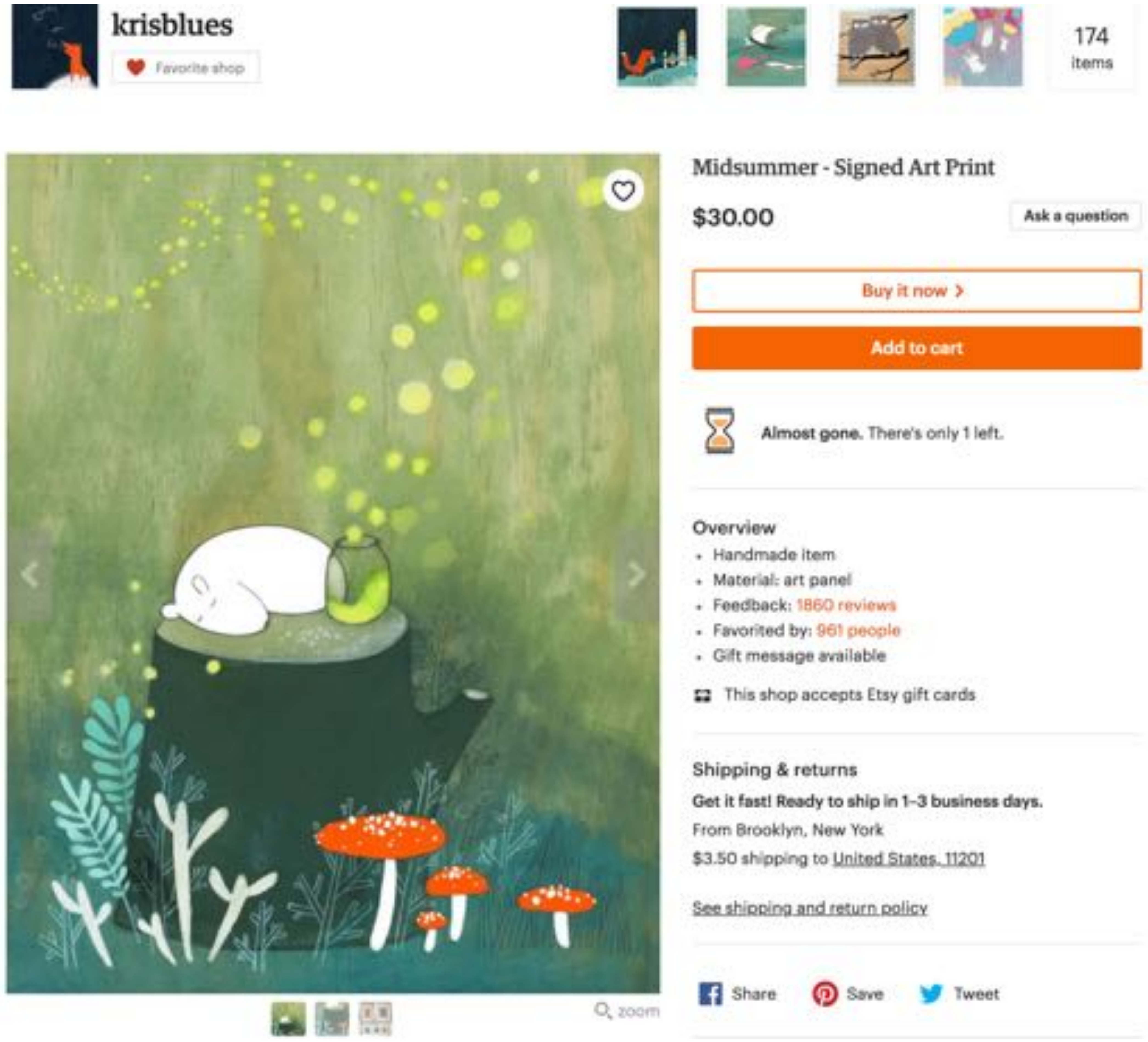
Challenges to Search for E-Commerce

Challenge II – User Satisfaction



Challenges to Search for E-Commerce

Challenge III – Discovery



Challenges to Search for E-Commerce

Challenges

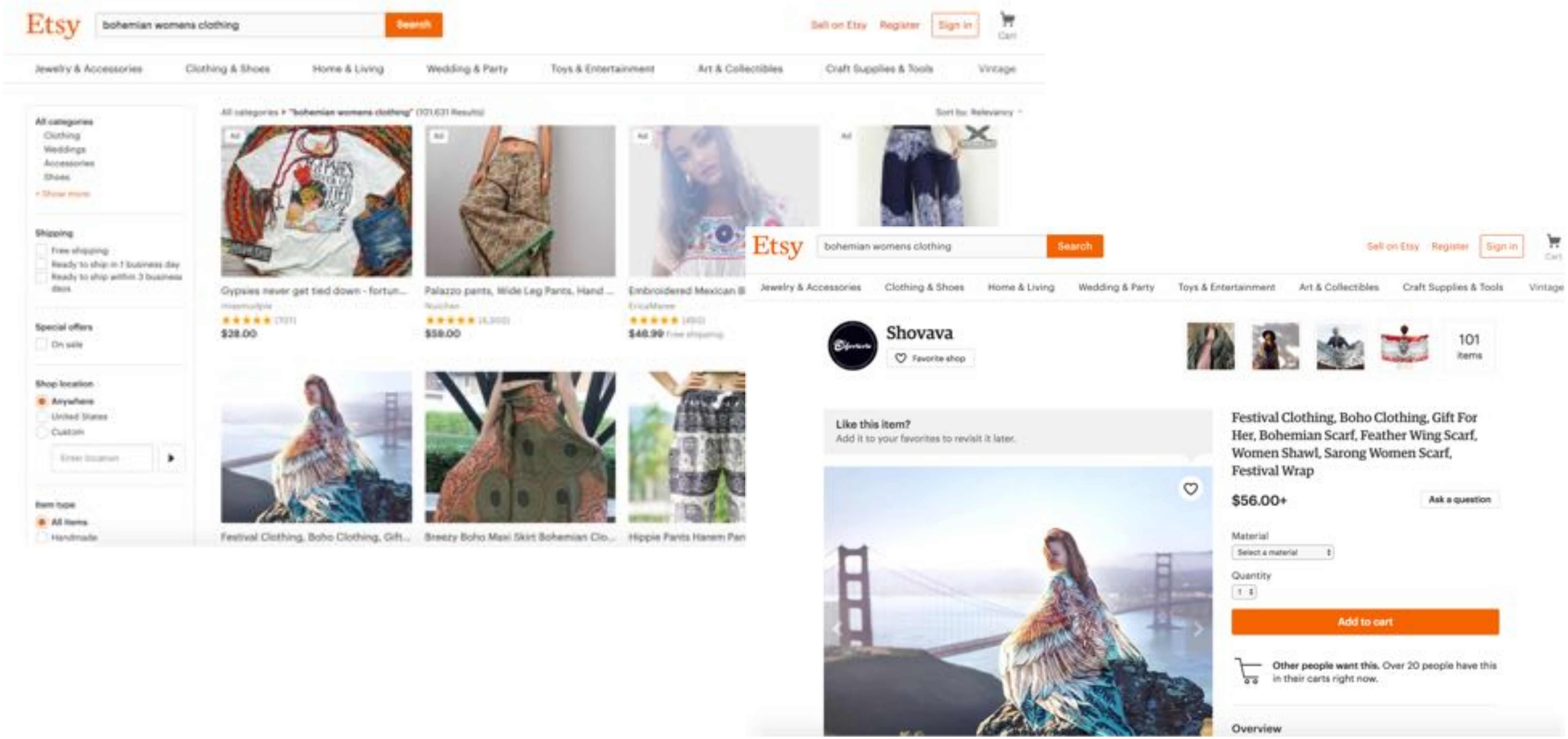
- Relevancy
- User Satisfaction
- Discovery

Publications

[1] Huizhong Duan, ChengXiang Zhai, Jinxing Cheng, and Abhishek Gattani. **Supporting Search in product database: a probabilistic approach**. In VLDB 2013.
[2*] Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. **Learning Latent Vector Spaces for Product Search**. In CIKM 2016.
[3] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. **On Application of Learning to Rank for E-Commerce Search**. In SIGIR 2017.
[4*] Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and W. Bruce Croft. **Learning a Hierarchical Embedding Model for Personalized Product Search**. In SIGIR 2017.
[5] Shichen Liu, Fei Xiao, Wenwu Ou, and Luo Si. **Cascade Ranking for Operational E-commerce Search**. In KDD 2017.
[6] Liang Wu, Diane Hu, Liangjie Hong, and Huan Liu. **Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce**. In SIGIR 2018.

Etsy’s Efforts on Search Ranking

Search Ranking at Etsy

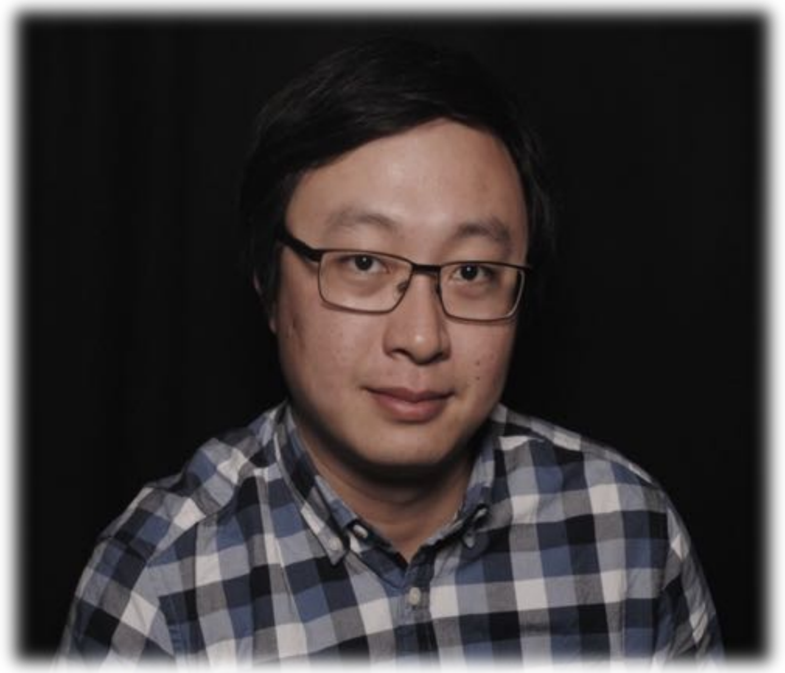


Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- **Liang Wu**, PhD Student from Arizona State University
- **Diane Hu**, Staff Data Scientist at Etsy
- **Liangjie Hong**, Head of Data Science at Etsy





Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

How to Optimize *Gross-Merchandise-Value* (GMV)?

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

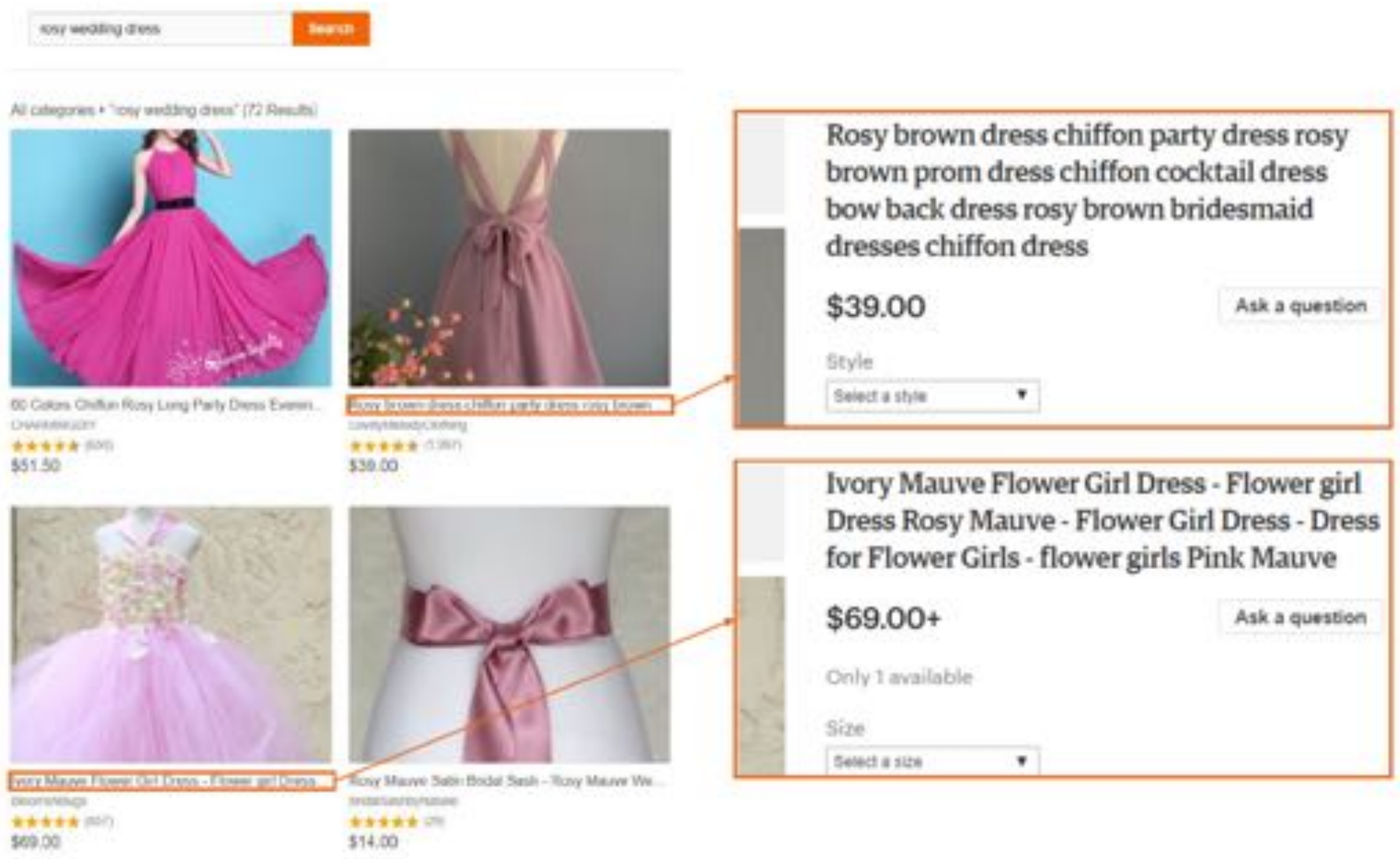
How to Optimize *Gross-Merchandise-Value* (GMV)?

$$GMV = \underbrace{\sum_{\forall s \in S}}_{\text{A search session}} \underbrace{\sum_{\forall i^s}}_{\text{An item in s}} \underbrace{Price(i^s)}_{\text{Price of } i^s} \underbrace{Pr(\Phi = 1 | i^s, q^s)}_{\text{Prob of purchase}},$$

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

Purchase Decision Process



Search Page Product Page

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- Click Decision(s) from Search-Result-Page (SERP)
- Purchase Decision(s) from Listing Page

$$Pr(\Phi = 1|i, q) = \underbrace{Pr(\Psi = 1|i, q)}_{\text{click model}} \underbrace{Pr(\Phi = 1|\Psi = 1, i, q)}_{\text{purchase model}},$$

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- Click Decision(s) from Search-Result-Page (SERP)
- Purchase Decision(s) from Listing Page

$$NDCG_K(\varrho) = N_{max}^{-1} \sum_{r=0}^{K-1} \frac{2^{l(r^{-1})}}{\log(1+r)},$$

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

- Click Decision(s) from Search-Result-Page (SERP)
- Purchase Decision(s) from Listing Page

$$NDCG_K(\varrho) = N_{max}^{-1} \sum_{r=0}^{K-1} \frac{2^{l(r^{-1})}}{\log(1+r)},$$

- ***l*** is transformed from *empirical GMV*.
- ***r*** is approximated by the product of a click model and a purchase model where the click model is a RankNet model and the purchase model is *price-weighted* logistic regression.

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

Relevance	Low Level	Sum of TF
		Sum of Log *TF*
		Sum of Normalized *TF*
		Sum of Log Normalized *TF*
		Sum of *IDF*
		Sum of Log *IDF*
		Sum of *ICF*
		Sum of *TF-IDF*
		Sum of Log *TF-IDF*
		TF-Log *IDF*
		Length
		Log *Length*
	High Level	*BM25*
		Log *BM25*
		LM_{DIR}
		LM_{JM}
		LM_{ABS}
Revenue		*Price*
		Price – Cat.Mean
		(Price – Cat.Mean)/Cat.Mean

Sessions	Queries	Items	Avg. Items per Session
334,931	239,928	6,347,251	19.0
Keywords	**Buyers**	**Sellers**	**Avg. Items per Query**
631,778	270,239	550,025	26.5

Click	RankNet [2]	RNet
	RankBoost [8]	RBoost
	AdaRank [36]	ARank
	LambdaRank [3]	LRank
	ListNet [4]	LNet
	MART [10]	MART
	LambdaMART [35]	LMART
Purchase	SVM [5]	SVM
	Logistic Regression [25]	LR
	Random Forest [19]	RM
Both	Weighted Purchase [41]	WT
	LMART+RM	LMRM
	LETORIF	LETORIF

Search Ranking at Etsy

Turning Clicks into Purchases: Revenue Optimization for Product Search in E-Commerce (SIGIR 2018)

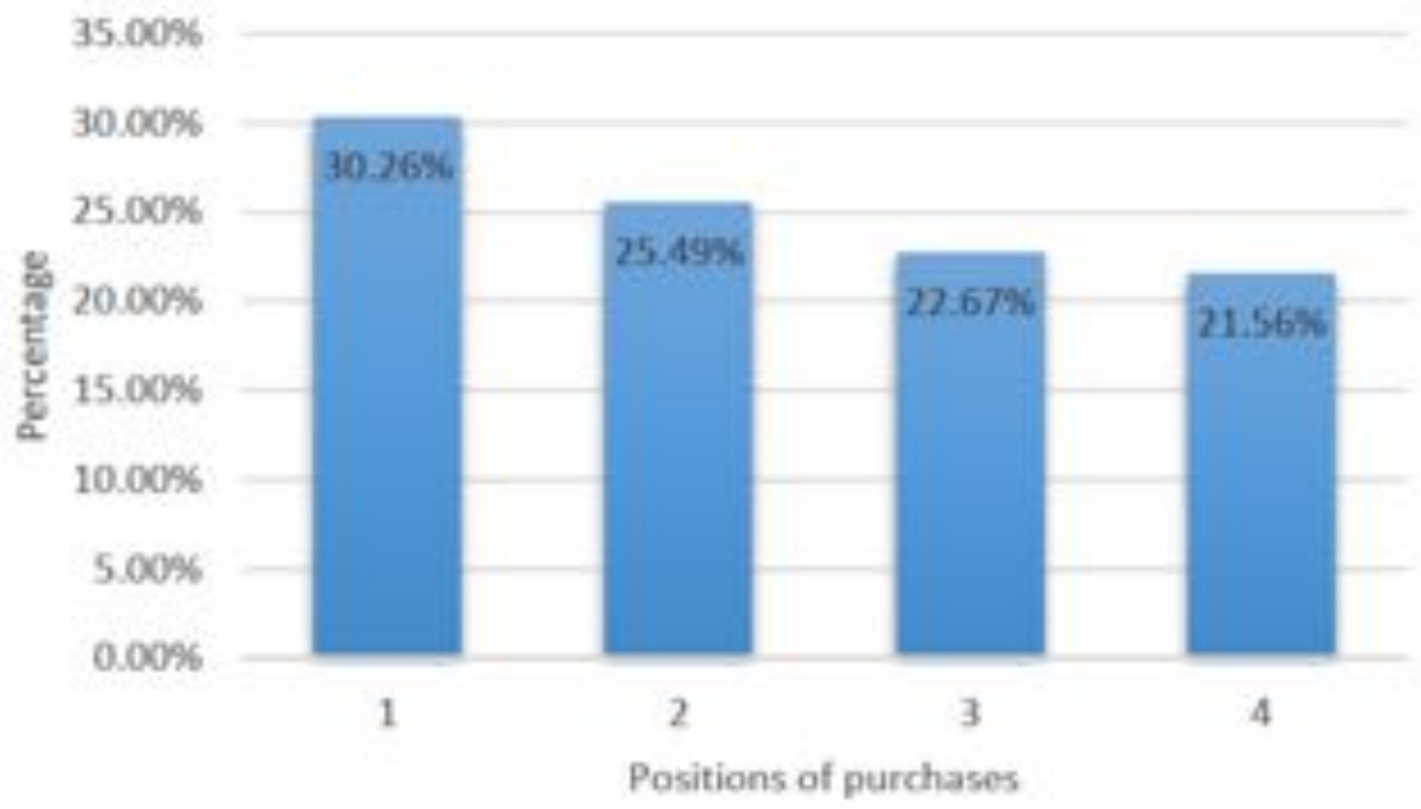


Figure 2: Position distribution of items being purchased in the top 4 spots of a search result page. The first position achieves the most purchases, while nearly 70% of purchases are in the lower positions.

Search Ranking at Etsy

Category	Method	Click NDCG@5			Purchase NDCG@5			Revenue NDCG@5		
		Train	Vali	Test	Train	Vali	Test	Train	Vali	Test
Click	RNet	0.1743	0.1731	0.1378**	0.1672	0.1721	0.1676**	0.1692	0.1700	0.1356**
	RBoost	0.2150	0.1768	0.1323**	0.2150	0.1768	0.1715**	0.2150	0.1768	0.1311**
	ARank	0.1718	0.1711	0.1351**	0.1718	0.1711	0.1706**	0.1718	0.1711	0.1358**
	LRank	0.1694	0.1688	0.1360**	0.1678	0.1711	0.1672**	0.1713	0.1719	0.1366**
	LNet	0.1665	0.1703	0.1355**	0.1601	0.1682	0.1620**	0.1646	0.1696	0.1348**
	MART	0.2700	0.1758	0.1380**	0.2155	0.1803	0.1796*	0.2696	0.1688	0.1408**
	LMART	0.3056	0.1777	**0.1412**	0.3056	0.1777	0.1717**	0.3056	0.1777	0.1370**
Purchase	SVM	0.1785	0.1772	0.1336**	0.1831	0.1754	0.1755**	0.1816	0.1752	0.1320**
	LR	0.1978	0.1739	0.1310**	0.1978	0.1739	0.1782**	0.1978	0.1739	0.1332**
	RM	0.3359	0.1698	0.1363**	0.3329	0.2305	0.1798**	0.3327	0.1685	0.1376**
Both	WT	0.1970	0.1682	0.1334**	0.1815	0.1763	0.1761**	0.1781	0.1648	0.1375**
	LMRM	0.2943	0.2597	0.1354**	0.3087	0.2530	0.1688**	0.2943	0.2594	0.1332**
	LETORIF	0.1765	0.1550	0.1351**	0.2731	0.1841	**0.1801**	0.2039	0.1698	**0.1494**

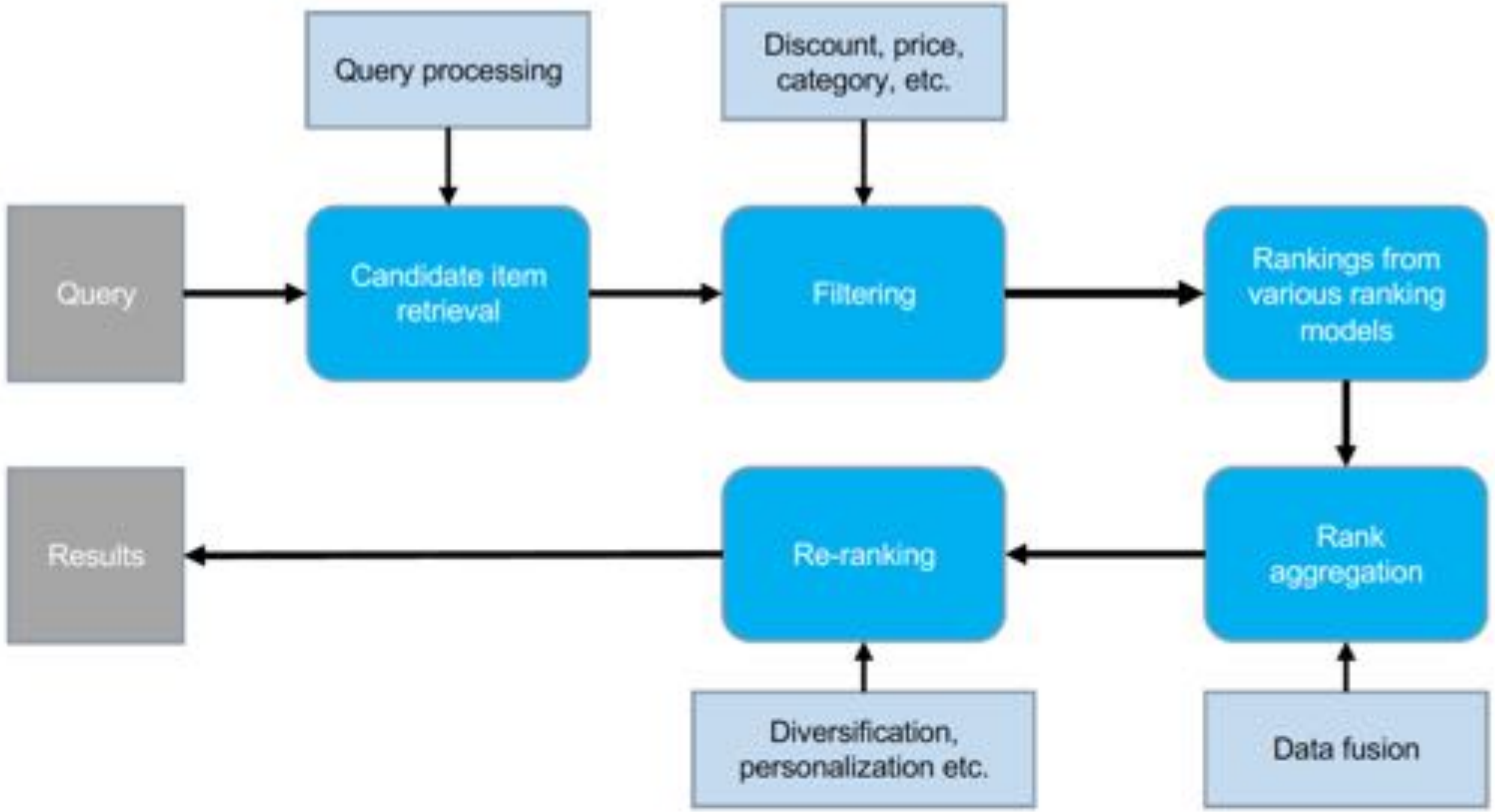
Symbol * indicates that the method is outperformed by the best one by 0.05 statistical significance level. ** indicates 0.01.

Search Ranking at Etsy

Category	Method	Rev@1	Rev@2	Rev@3	Rev@4	Rev@5	Rev@6	Rev@7	Rev@8	Rev@9	Rev@10
Click	RNet	4.47**	4.69**	4.89**	4.91*	5.06**	5.23**	5.21**	5.33**	5.46**	5.55**
	RBoost	4.57**	4.69**	4.69**	4.76**	4.97**	5.17**	5.23**	5.36**	5.49**	5.57**
	ARank	4.37**	4.66**	4.76**	4.90**	5.06**	5.20*	5.33**	5.47**	5.59**	5.67**
	LRank	4.38**	4.61**	4.74**	4.86**	5.07**	5.25**	5.42**	5.42**	5.67**	5.78**
	LNet	4.30**	4.59**	4.78**	4.99**	5.16**	5.35**	5.49**	5.61**	5.63**	5.63**
	MART	**4.62**	4.72**	4.86**	5.04**	5.26**	5.47**	5.47**	5.64**	5.74**	5.86**
	LMART	4.46*	4.54**	4.73**	5.10**	5.31**	5.56**	5.75**	5.90*	6.01**	6.14**
Purchase	SVM	4.41**	4.54**	4.76**	4.77**	4.95**	5.16**	5.34**	5.50**	5.64**	5.77**
	LR	4.29**	4.65**	4.65**	4.69**	4.74**	4.81*	4.94**	4.97**	5.11**	5.11**
	RM	4.52**	4.82**	4.86**	5.02**	5.18**	5.33*	5.50**	5.66**	5.79**	5.92**
Both	WT	4.52**	4.69**	4.80**	4.85**	5.01**	5.07**	5.23**	5.32**	5.35**	5.41**
	LMRM	4.42**	4.50**	4.72**	5.08**	5.23**	5.41**	5.57**	5.60**	5.73**	5.85**
	LETORIF	4.58**	**4.90**	**5.08**	**5.47**	**5.64**	**5.85**	**6.02**	**6.19**	**6.40**	**6.54**

Symbol * indicates that the method is outperformed by the best one by 0.05 statistical significance level, ** indicates 0.01.

Search Ranking at Etsy



Credit: SIGIR 2018 Tutorial Information Discovery in E-Commerce

Search Ranking at Etsy

- A simplified 2-Stage model deployed into recommendation, improved GMV +0.8%.
- A weighted purchase model deployed into search ranking, improved GMV +0.9%.
- An extended candidate selection deployed into search ranking, improved GMV +2.0%.
- A model heavily utilizing historical information deployed into search ranking, improved GMV +0.7%.

Conclusion

- **An Introduction to Etsy**
- **Challenges to Search for E-Commerce**
- **Etsy’s Efforts on Search Ranking**

Conclusion

- **An Introduction to Etsy**
- **Challenges to Search for E-Commerce**
- **Etsy's Efforts on Search Ranking**
- **Incoming Kaggle Competition**

Thanks!