

Crowdsourcing Backstories for Complex Task-Based Search

Manuel Steiner, Damiano Spina, Falk Scholer and Lawrence Cavedon

3 February 2022

RMIT University, Melbourne, Australia

Job search query: english teacher

Job search query: english teacher

Backstory 1:

You are an **English teacher** that just moved to the **Eastern Suburbs of Melbourne** and are looking for a new **full time** job. You have experience teaching VCE English as well as Year 7 English and Year 8 Humanities.

Job search query: english teacher

Backstory 1:

You are an **English teacher** that just moved to the **Eastern Suburbs of Melbourne** and are looking for a new **full time** job. You have experience teaching VCE English as well as Year 7 English and Year 8 Humanities.

Backstory 2:

You're a recent graduate but do have student teaching experience.

You're looking for an **entry level English teaching** position to get started on your career. You're **flexible about location**.

- Information need statements
- Relevance judgements, query disambiguation
- Available for popular IR settings

- Information need statements
- Relevance judgements, query disambiguation
- Available for popular IR settings
- Crowdsourcing backstories
- Cost-effective, variety
- Methodology, job search collection

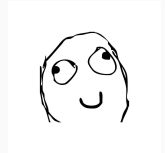
Personas

- Elaborate system user descriptions
- Additional artefacts, info
- Scenarios comparable to backstories

User Interaction Modelling

Personas

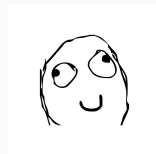
- Elaborate system user descriptions
- Additional artefacts, info
- Scenarios comparable to backstories



User Interaction Modelling

Personas

- Elaborate system user descriptions
- Additional artefacts, info
- Scenarios comparable to backstories



Backstories

- Short statement
- Single information need
- Enough info for task

Crowdsourcing Backstories - Task I

Search Query
manager

Job Advertisement

Location: Melbourne

Work type: Full Time

Classification: Management

Salary: No data available

We are looking for a new manager!

Figure 1: The final interface of the crowdsourcing task (I).

Crowdsourcing Backstories - Task II

Job advertisement relevance

Is the job advertisement relevant to the search query?

- ☐ Yes, the job advertisement is a relevant result for the search query.
- ☐ No, the job advertisement is not a relevant result for the search query.

Backstory generation

Write a backstory that would make the job advertisement satisfactory to the given search request. **Imagine the situation of a job seeker! Use the English language in second form ('You are...', 'You do...').**

What is the current background situation of the job seeker?

How satisfactory is the job advertisement in relation to the backstory that you wrote?

	1	2	3	4	5	
not satisfactory at all	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	very satisfactory

Please explain your satisfaction decision.

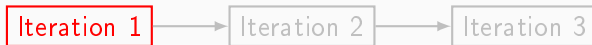
Why did you choose the above rating for your backstory?

Figure 2: The final interface of the crowdsourcing task (II).

Crowdsourcing Backstories - Setup

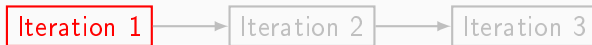
- 15 queries and ads, 5 backstories each
- Major job search platform, May - June 2017
- Platform: Figure-Eight (CrowdFlower)
- Workers from English speaking countries

Crowdsourcing Backstories - First Iteration



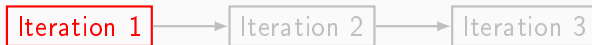
- Minimum 100 characters
- Minimum 3 minutes
- US\$0.5 payment
- No accuracy checks, worker filtering

Crowdsourcing Backstories - First Iteration

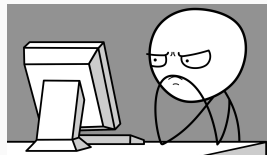


- Minimum 100 characters
- Minimum 3 minutes
- US\$0.5 payment
- No accuracy checks, worker filtering
- 70% usable backstories
- Hirer's perspective, copy and paste

Crowdsourcing Backstories - First Iteration



- Minimum 100 characters
- Minimum 3 minutes
- US\$0.5 payment
- No accuracy checks, worker filtering
- 70% usable backstories
- Hirer's perspective, copy and paste



Crowdsourcing Backstories - Second Iteration



- Improved instructions

Crowdsourcing Backstories - Second Iteration



- Improved instructions
- Prohibit copy and paste into text areas
- Substring detection of job ad in text areas

Crowdsourcing Backstories - Second Iteration



- Improved instructions
- Prohibit copy and paste into text areas
- Substring detection of job ad in text areas
- US\$0.65 payment

Crowdsourcing Backstories - Second Iteration



- Improved instructions
- Prohibit copy and paste into text areas
- Substring detection of job ad in text areas
- US\$0.65 payment
- 63% usable backstories
- Meaningless (template)
- 11 accounts (single person, info sharing)

Crowdsourcing Backstories - Second Iteration



- Improved instructions
- Prohibit copy and paste into text areas
- Substring detection of job ad in text areas
- US\$0.65 payment
- 63% usable backstories
- Meaningless (template)
- 11 accounts (single person, info sharing)



Crowdsourcing Backstories - Third Iteration



- Gold rows for accuracy checks (query relevance to job ad)
- Two tasks per page
- Initial quiz mode to filter workers

Crowdsourcing Backstories - Third Iteration



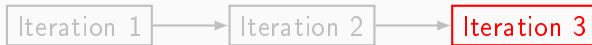
- Gold rows for accuracy checks (query relevance to job ad)
- Two tasks per page
- Initial quiz mode to filter workers
- Larger worker pool with lower confidence level
- Levenshtein similarity of backstories < 0.65

Crowdsourcing Backstories - Third Iteration



- Gold rows for accuracy checks (query relevance to job ad)
- Two tasks per page
- Initial quiz mode to filter workers
- Larger worker pool with lower confidence level
- Levenshtein similarity of backstories < 0.65
- minimum 200 characters

Crowdsourcing Backstories - Third Iteration



- Gold rows for accuracy checks (query relevance to job ad)
- Two tasks per page
- Initial quiz mode to filter workers
- Larger worker pool with lower confidence level
- Levenshtein similarity of backstories < 0.65
- minimum 200 characters
- 91% usable backstories
- Hirer's perspective

Crowdsourcing Backstories - Third Iteration



- Gold rows for accuracy checks (query relevance to job ad)
- Two tasks per page
- Initial quiz mode to filter workers
- Larger worker pool with lower confidence level
- Levenshtein similarity of backstories < 0.65
- minimum 200 characters
- 91% usable backstories
- Hirer's perspective



Analysis of Backstories

Table 1: Example backstories that were obtained via the third iteration.

<i>Query</i>	<i>Crowdsourced Backstory</i>
mining sales	You have 10 years experience in the sales sector including 5 years as management. You are very effective at spurring growth and have been working in the mining sector for the past two of those years. Due to relocation you would like to find a sales position in the Perth area.
welder	You are a certified welder, familiar with every tool and are looking for work in a factory environment. You have over 7 years experience working in a factory and are a capable machine operator as well.

Comparison of Crowdsourcing Iterations I

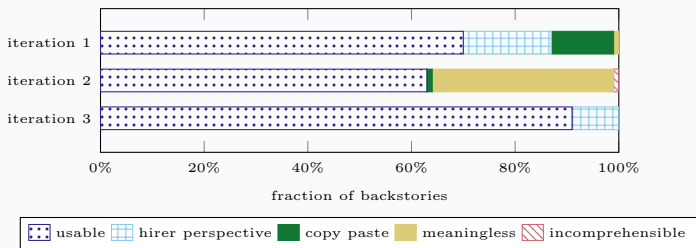


Figure 3: Fractions of backstories per type and iteration.

Comparison of Crowdsourcing Iterations II

Table 2: Descriptive statistics of the three iterations. The cost of iteration three include tasks that were used for accuracy checks. The backstory lengths are based on usable backstories.

Iter.	Total Cost (US\$)	Usable Stories (%)	Overall Completion Time (h)	Mean Task Completion Time	# Workers	Stories per Worker			Story Length (words)		
						Min	Median	Max	Min	Median	Max
1	45	70	6	4m55s	21	1	2	10	16	30	62
2	68	61	4	5m04s	31	1	2	10	16	32	76
3	171	91	85	5m27s	68	1	2	10	33	50	149

Table 3: Descriptive statistics of the job search backstory collection.

Number of queries	50
Number of backstories	756
Minimum number of backstories per query	10
Average number of backstories per query	15
Maximum number of backstories per query	28
Minimum number of backstories per worker	2
Median number of backstories per worker	4
Maximum number of backstories per worker	40
Minimum task completion time	1m45s
Average task completion time	6m40s
Maximum task completion time	15m0s
Minimum backstory length (words)	8
Median backstory length	45
Maximum backstory length	234

Conclusions

- Backstory crowdsourcing methodology
- 91% usable backstoies
- Potentially transferable
- Similar domains - real estate, automobile sales

Conclusions

- Backstory crowdsourcing methodology
- 91% usable backstoies
- Potentially transferable
- Similar domains - real estate, automobile sales
- Job search backstory collection
- 756 backstories in 5 months

Conclusions

- Backstory crowdsourcing methodology
- 91% usable backstoies
- Potentially transferable
- Similar domains - real estate, automobile sales
- Job search backstory collection
- 756 backstories in 5 months
- Small scale test crowdsourcing setups
- For more complex tasks

Future Work

- Backstory usability, quality assessment
- Other task-based search settings

Future Work

- Backstory usability, quality assessment
- Other task-based search settings

Resources

<https://github.com/manuelsteiner/job-search-backstories>

Future Work

- Backstory usability, quality assessment
- Other task-based search settings

Resources

<https://github.com/manuelsteiner/job-search-backstories>

Acknowledgements

- Australian Research Council Linkage Project LP150100252
- SEEK Ltd.

My Backstory

Manuel's Backstory:

I am in the finishing stages of my PhD researching backstories in search. I am looking for employment opportunities. My expertise and interests include information retrieval, data analytics and manipulation, software development and related areas.

My Backstory

Manuel's Backstory:

I am in the finishing stages of my PhD researching backstories in search. I am looking for employment opportunities. My expertise and interests include information retrieval, data analytics and manipulation, software development and related areas.

Contact

- Email - mail@linkto.ms
- Twitter - [@manuelsteiner_](https://twitter.com/manuelsteiner_)