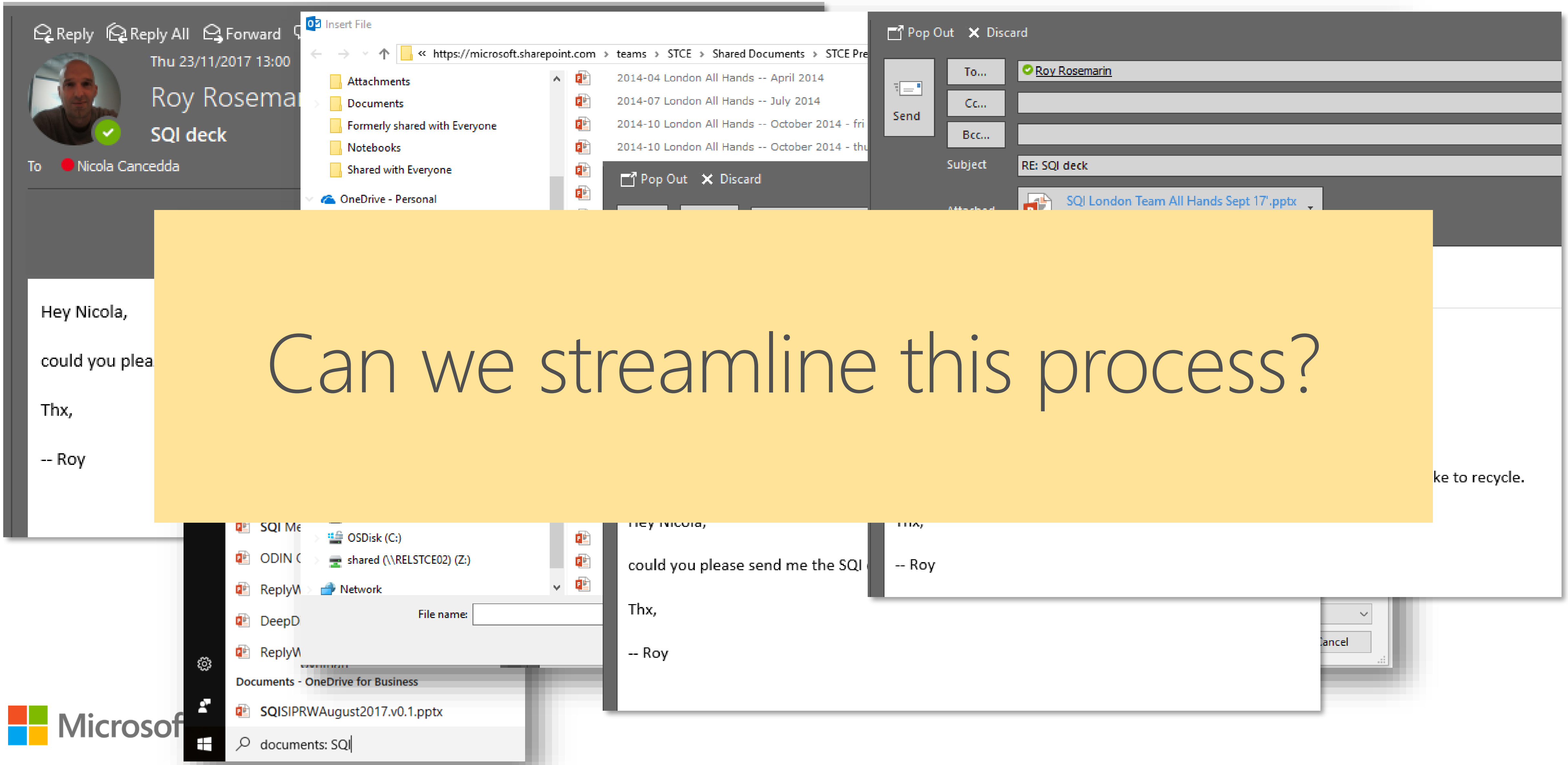
Reply With: Suggesting Email Attachments

Nicola Cancedda

Microsoft

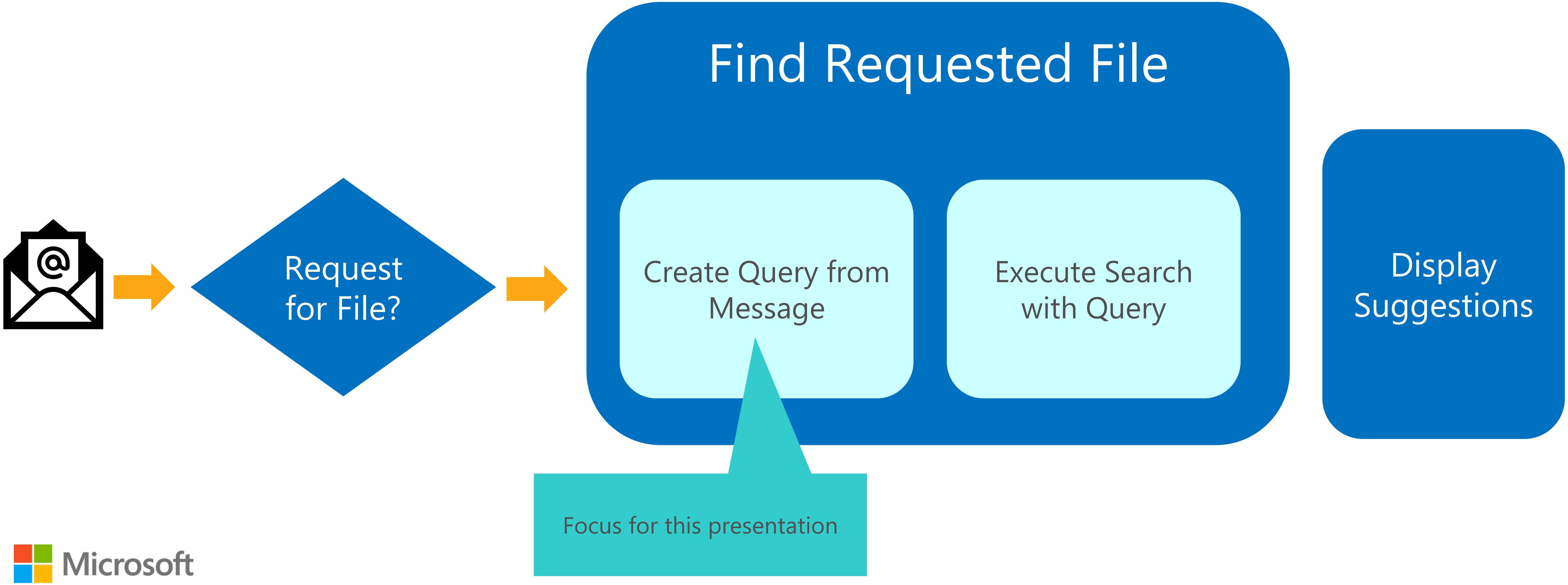
November
2017

Responding to a File Request



Can we streamline this process?

Suggesting Content to Share – Decomposing the Problem

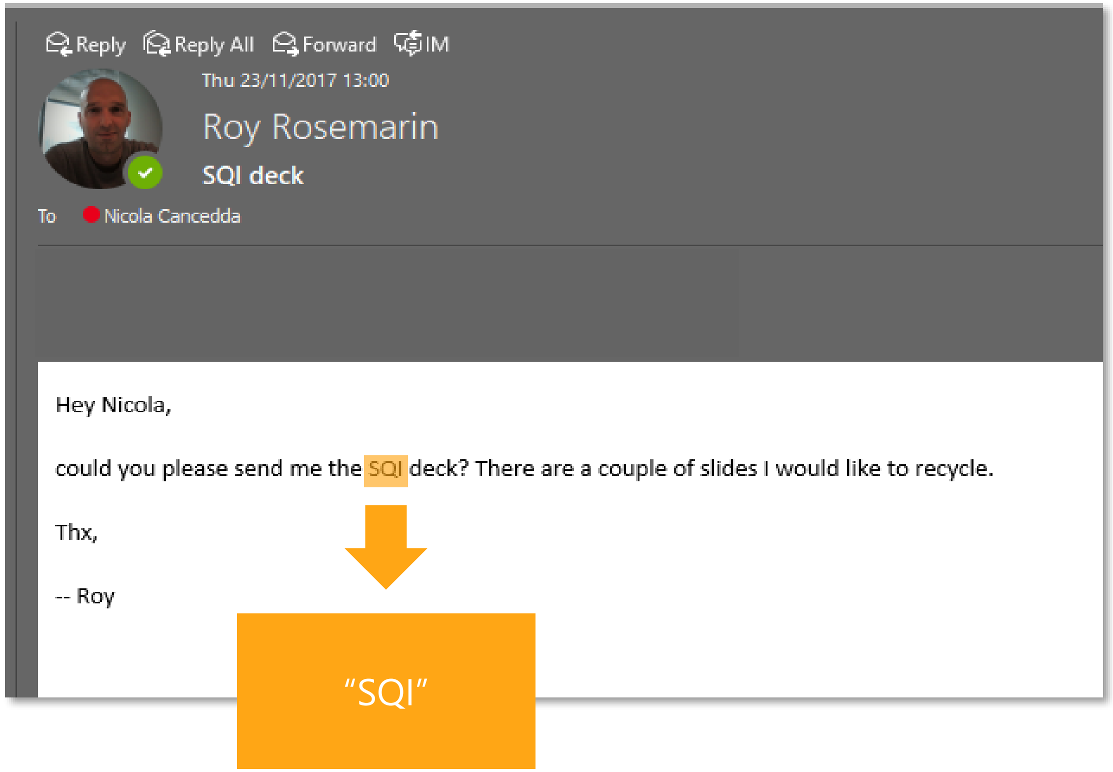


Learning to Create Queries from Email Messages

- Problem: Query Term Extraction and Ranking

- Related work:
  - Query construction for searching for Prior Art from Patent Applications [Xue and Croft, 2009; Cetintas and Si, 2012]
  - Verbose query simplification [Bendersky and Croft, 2008, Xue et al. 2010]
  - Keyword extraction from documents

- Approach:
  - Generate an initial set of candidates with high TF-IDF
  - Train a discriminative model to rerank candidate terms

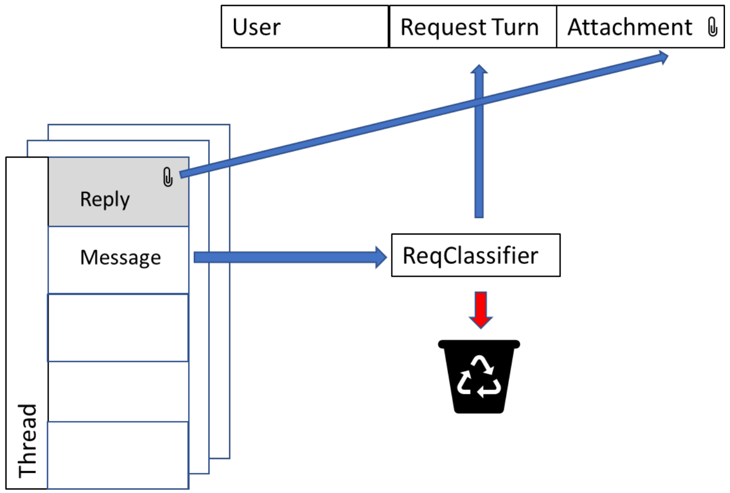


Labelled Data Anyone? Distant Supervision

We need request emails paired with "good queries" to train and evaluate Query Ranking components

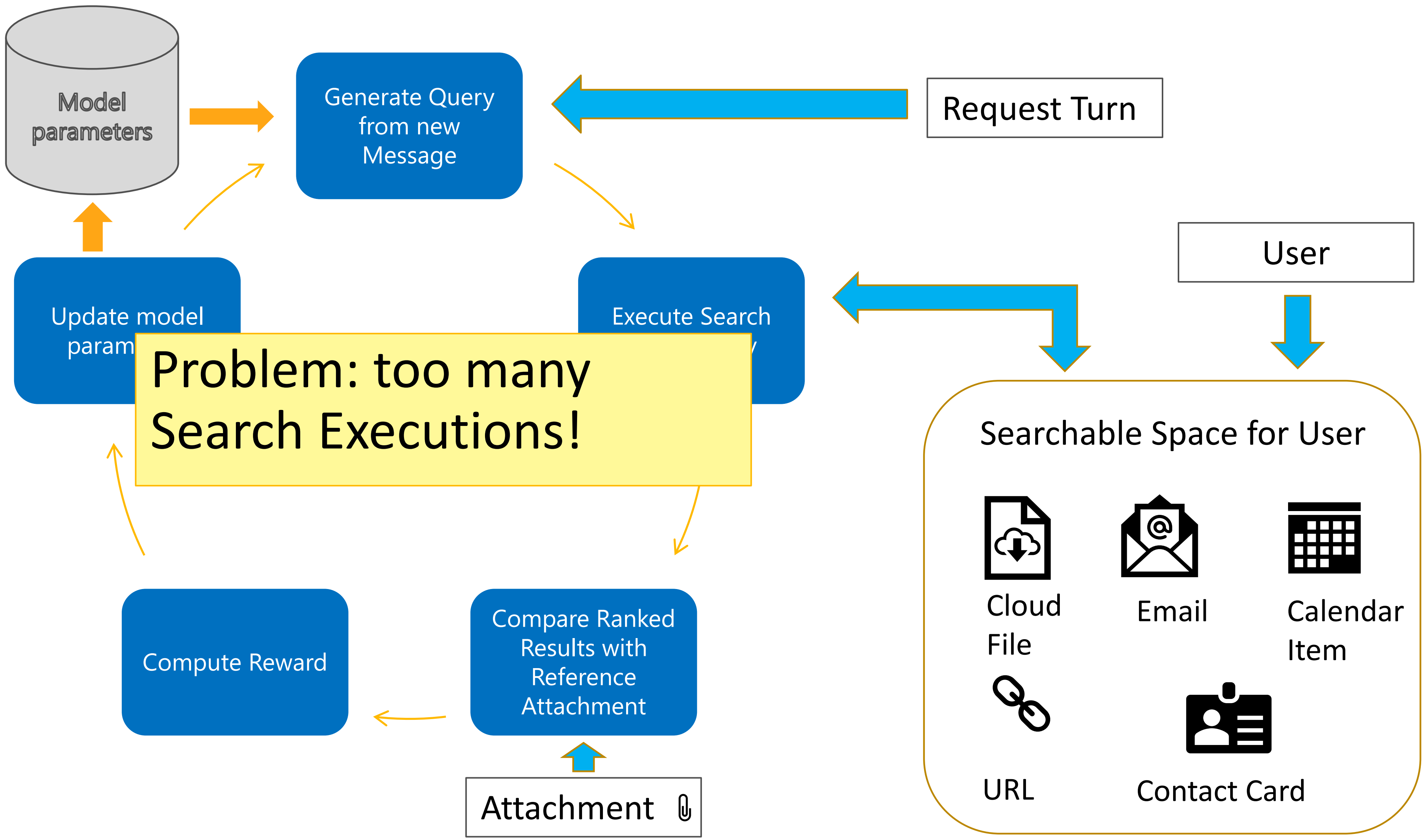
There is no such thing as a good query **in absolute**.

Query quality depends inherently on the Search engine it is sent to. A query is good if, **when issued to a given Search engine**, it retrieves the desired entity "close to the top"



Microsoft

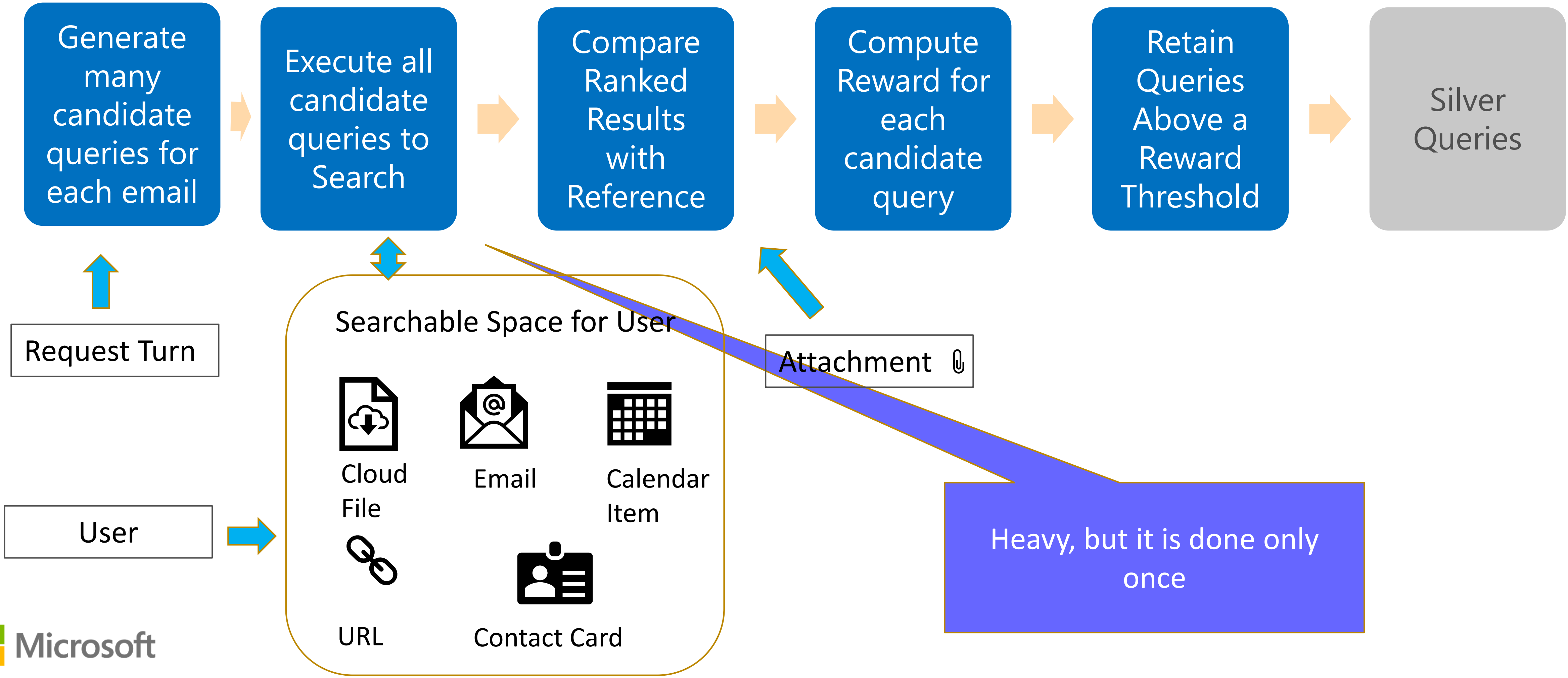
Query Ranker Training - Ideal M2Q Learning Cycle



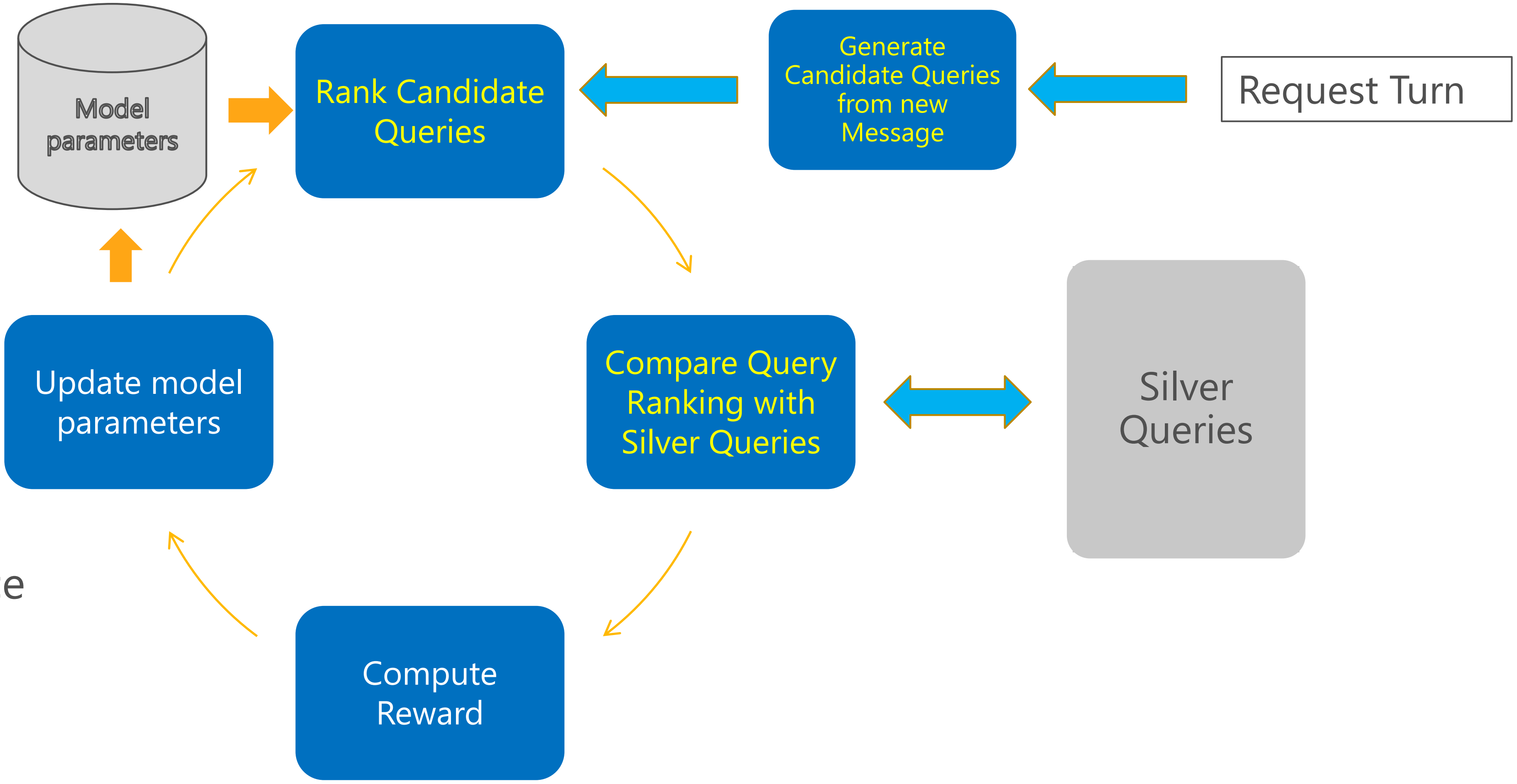
Microsoft

Query Ranker Training

Approximate with "Silver Queries". Upfront, before training begins:



Query Ranker Training - M2Q as Learning to Rank



- Slightly Weaker Guidance from Silver Queries

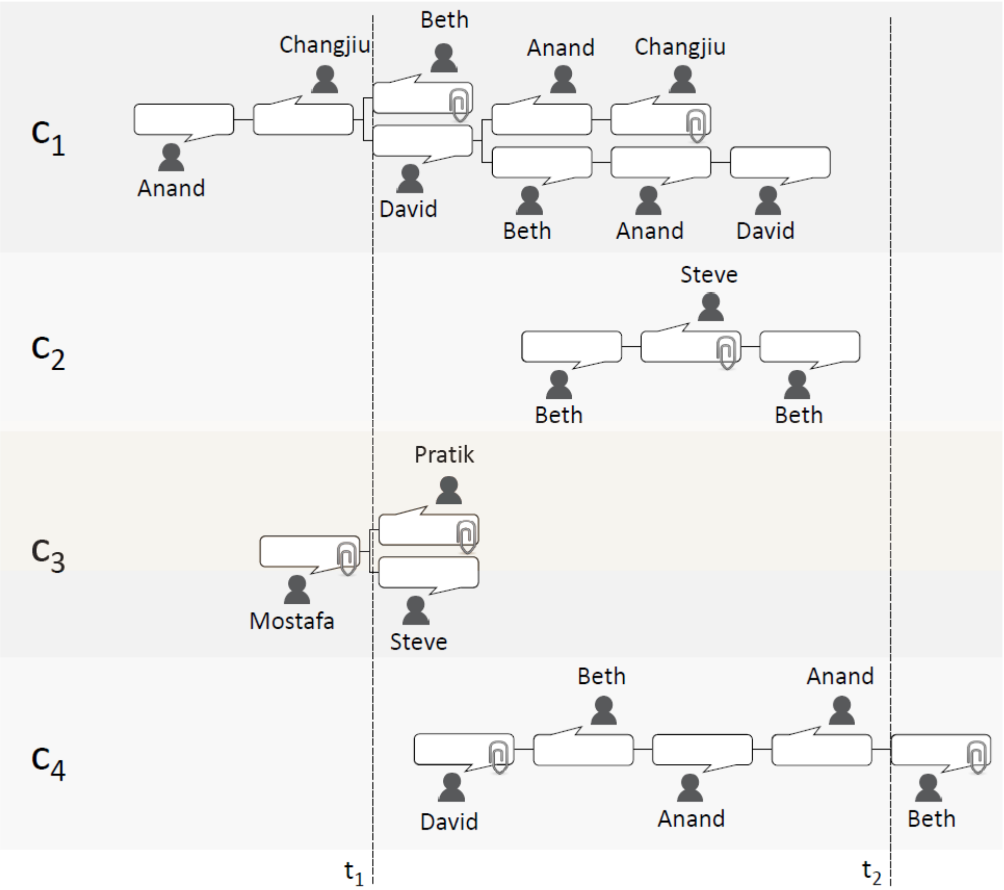
+ Much faster!

Proxy Problem

Given an incoming request message, find a previous message in the same mailbox that has the relevant item attached (if file) or included inline (if URL).

35-40% of attachments/inline mentions are already present in the sender mailbox.

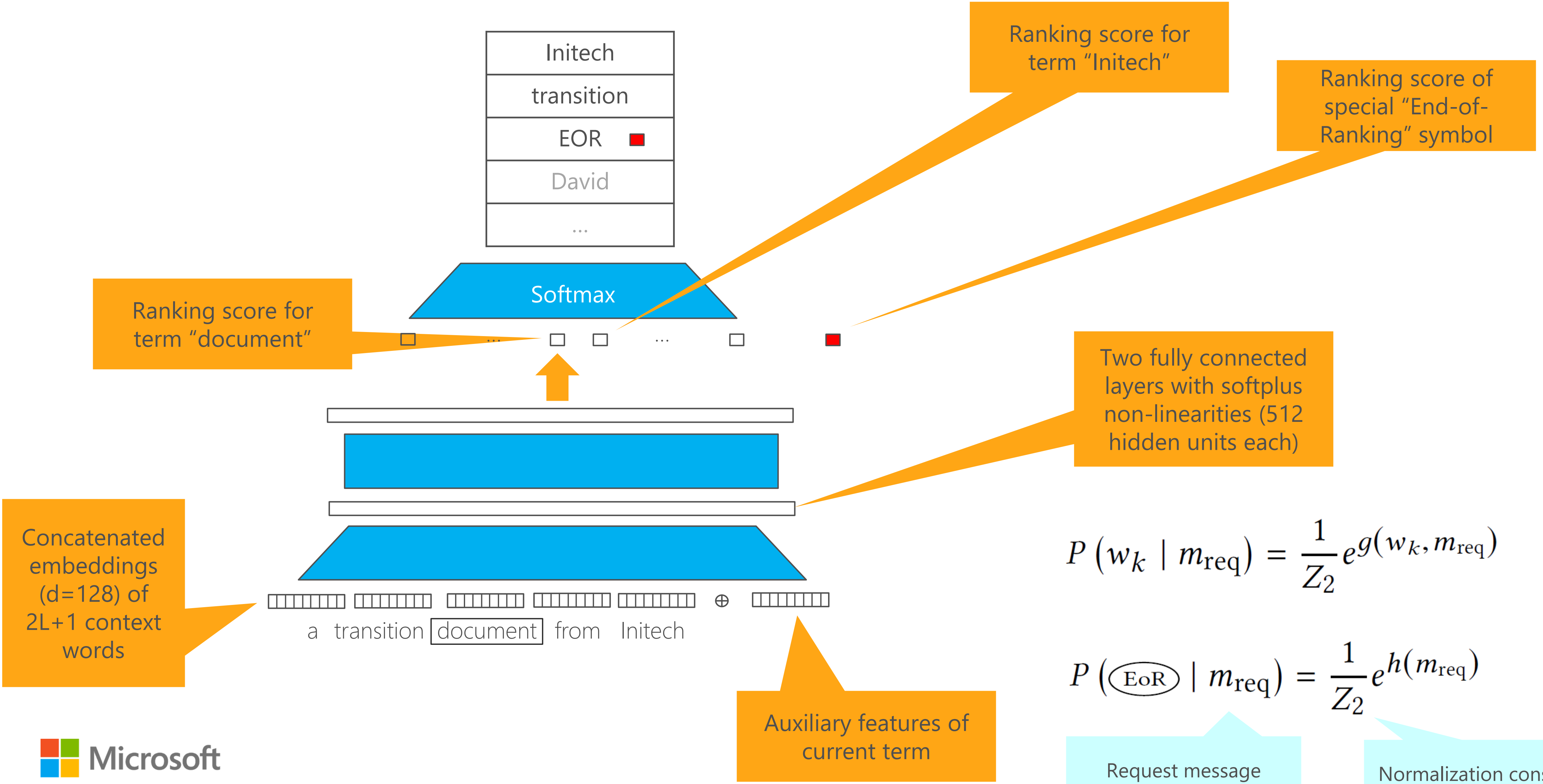
Beth's mailbox



From: Van Gysel et al. 2017

Microsoft

CNN Architecture



$$P\left(w_k \mid m_{\text{req}}\right) = \frac{1}{Z_2} e^{g(w_k, m_{\text{req}})}$$

$$P\left(\text{EoR} \mid m_{\text{req}}\right) = \frac{1}{Z_2} e^{h(m_{\text{req}})}$$

Request message

Normalization constant

Model Features

- Word Embeddings:
  - Learned during training
  - Alternatively: pre-initialized using GloVe
- Auxiliary features:
  - POS features
  - Message features
  - Collection Statistics

Context features (learned representations)	
`term`	Representation of the term.
`context`	Representations of the context surrounding the term.
Part-of-Speech features	
`is_noun`	POS tagged as a noun [6]
`is_verb`	POS tagged as a verb
`is_other`	POS tagged as neither a noun or a verb
Message features	
`is_subject`	Term occurrence is part of the subject [14]
`is_body`	Term occurrence is part of the body [14]
`Abs. TF`	Abs. term freq. within the message [63]
`Rel. TF`	Rel. term freq. within the message [63]
`Rel. pos.`	Rel. position of the term within the message
`is_oov_repr`	Term does not have a learned representation
Collection statistics features	
`IDF`	Inverse document frequency of the term [63]
`TF-IDF`	$\texttt{TF} \times \texttt{IDF}$ [63]
`Abs. CF`	Abs. collection freq. within the collection
`Rel. CF`	Rel. collection freq. within the collection
`Rel. Entropy`	KL divergence from the unsmoothed collection term distribution to the smoothed ($\lambda = 0.5$) document term distribution [37]
`SCQ`	Similarity Collection/Query [68]
`ICTF`	Inverse Collection Term Frequency [31]
`Pointwise SCS`	Pointwise Simplified Clarity Score [24]

Microsoft

From: Van Gysel et al. 2017

Training Loss

$$L_{\text{xent}}(\theta \mid m_{\text{req}}, \tilde{q}) = -\sum_{\omega \in \Omega} Q(\omega \mid \tilde{q}) \log(P(\omega \mid m_{\text{req}}))$$

$$\Omega = (w_1, \ldots, w_n, \text{EoR})$$

First Loss term: cross-entropy with reference distribution

$$Q(w_k \mid \tilde{q}) = \alpha \cdot \frac{\mathbb{1}_{\tilde{q}}(w_k)}{\#(w_k, m_{\text{req}}) \cdot |\tilde{q}|}$$

$$Q(\text{EoR} \mid \tilde{q}) = (1 - \alpha)$$

For each training sample, the reference distribution is uniform over all distinct terms in the Silver Query, leaving out a constant mass for the EOR symbol

$$L_{\text{cutoff}}(\theta \mid m_{\text{req}}, \tilde{q}) = \left(\min_{w \in \tilde{q}}(g(w, m_{\text{req}})) - h(m_{\text{req}})\right)^2$$

Second Loss term: calibration of EOR score

$$L(\theta \mid B) = \frac{1}{|B|} \sum_{(m, \tilde{q}) \in B} \text{score}(\tilde{q}) \left(L_{\text{xent}}(\theta \mid m, \tilde{q}) + L_{\text{cutoff}}(\theta \mid m, \tilde{q})\right) + \frac{1}{2\lambda} \sum_{W \in \theta_W} \sum_{ij} W_{ij}^2$$

Regularization term

Batch uniformly sampled.

Samples weighted by score of Silver Query

From: Van Gysel et al. 2017

Experimental Data and Setup

- Data:
  - Avocado:
    - The Avocado collection is a public data set that consists of emails taken from 279 custodians of a defunct information technology company
  - PIE:
    - Internal Microsoft email obtained through an employee participation program

- Cross-validation
  - Train on corpus A and Test on corpus B
  - 95/5 Train/Development split

- Search:
  - Indri (https://www.lemurproject.org/indri/)
  - Query Likelihood Model with Dirichlet smoothing

Files or URLs

	Avocado	PIE
Messages	928,992	1,047,311
Message length (terms)	112.33 ± 244.01	74.70 ± 551.88
Threads	804,010	381,448
Thread lengths	1.19 ± 0.70	2.75 ± 3.65
Time period	3 years, 8 months	1 year
Attachable entities	50,462	28,725
Impressions per item	3.48 ± 2.55	2.79 ± 1.36
Messages with an item	311,478	152,649
no thread history	288,099	69,796
all items filtered (§5.3)	22,399	80,717
Request/reply pairs	980	2136
Thread history length of pairs	1.53 ± 1.13	4.04 ± 5.78
Relevant items per pair	1.22 ± 0.70	1.29 ± 1.82

From: Van Gysel et al. 2017

Experimental Results

Using the whole subject is a strong baseline

	Avocado MRR	Avocado NDCG	Avocado P@5	PIE MRR	PIE NDCG	PIE P@5
Full field, single features and random (subject)						
Full	0.2286	0.3097	0.0686	0.3338	0.4621	0.1088
TF	0.2280	0.3095	0.0686	0.3315	0.4600	0.1079
TF-IDF	0.2250	0.3073	0.0704	0.3390	0.4663	0.1090
logTF-IDF	0.2280	0.3095	0.0686	0.3315	0.4600	0.1079
RE	0.2223	0.3038	0.0698	*0.3391*	*0.4664*	*0.1095*
Random *k*	0.2143	0.2932	0.0647	0.3266	0.4553	0.1063
Random %	0.1481	0.2104	0.0467	0.2749	0.4013	0.0889

	Avocado MRR	Avocado NDCG	Avocado P@5	PIE MRR	PIE NDCG	PIE P@5
Full field, single features and random (body)						
Full	0.1248	0.1930	0.0377	0.2115	0.3376	0.0672
TF	0.1025	0.1719	0.0309	0.2094	0.3358	0.0660
TF-IDF	0.1507	0.2213	0.0459	0.2237	0.3481	0.0722
logTF-IDF	0.1109	0.1755	0.0311	0.1914	0.3180	0.0627
RE	0.1441	0.2128	0.0424	0.2198	0.3430	0.0699
Random *k*	0.0785	0.1394	0.0229	0.1781	0.3078	0.0568
Random %	0.1030	0.1646	0.0325	0.1887	0.3128	0.0606

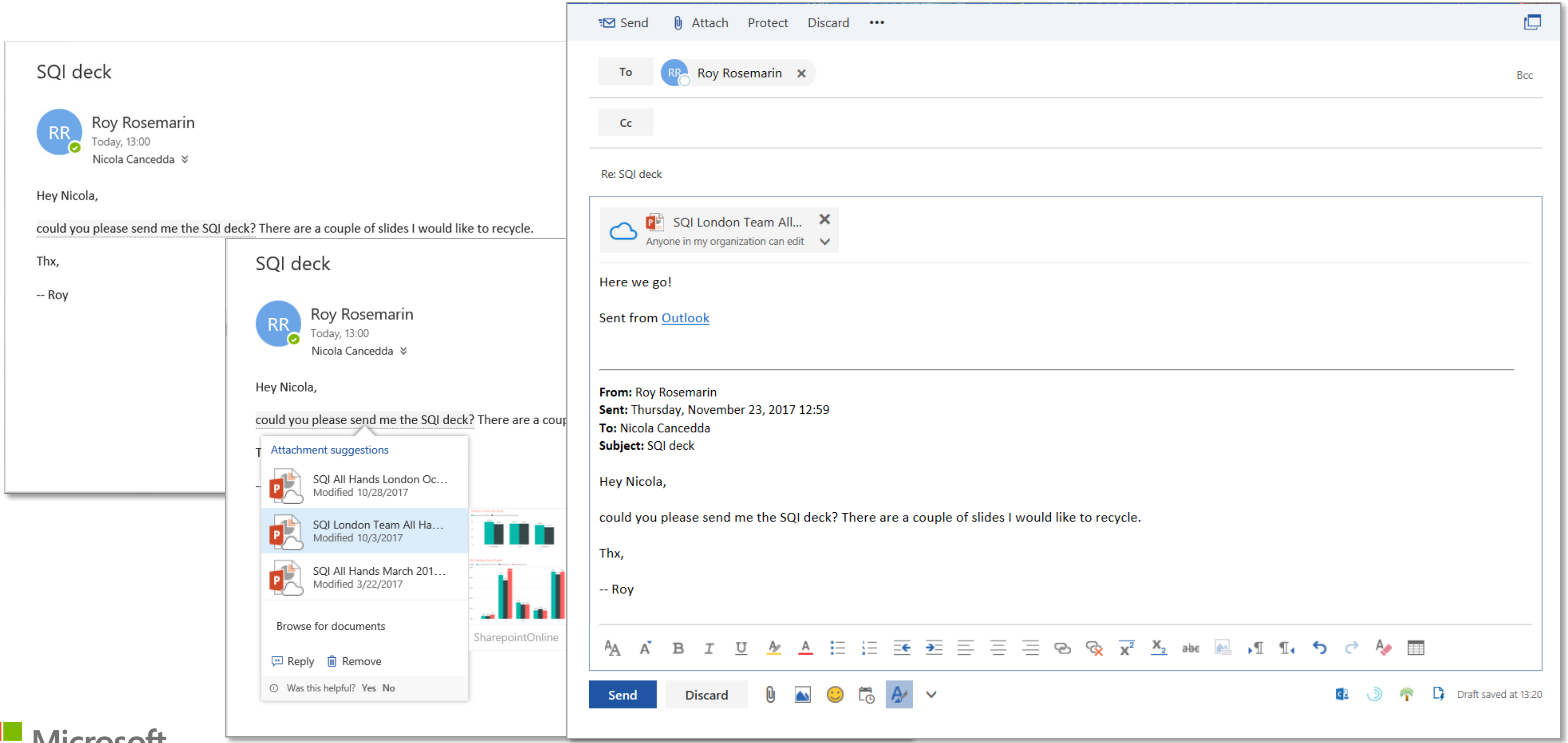
	Avocado MRR	Avocado NDCG	Avocado P@5	PIE MRR	PIE NDCG	PIE P@5
Full field, single features and random (subject + body)						
Full	0.1995	0.2785	0.0612	0.3087	0.4406	0.0972
TF	0.1783	0.2653	0.0551	0.3005	0.4334	0.0953
TF-IDF	0.2097	0.2933	0.0649	0.3100	0.4397	0.0991
logTF-IDF	0.1858	0.2726	0.0592	0.2747	0.4098	0.0871
RE	0.2138	0.2980	0.0649	0.3200	0.4489	0.1023
Random *k*	0.1404	0.2148	0.0436	0.2721	0.4076	0.0886
Random %	0.1753	0.2514	0.0520	0.2592	0.3941	0.0822

	Avocado MRR	Avocado NDCG	Avocado P@5	PIE MRR	PIE NDCG	PIE P@5
Learning-to-rank methods (subject + body)						
RankSVM	0.1650	0.2425	0.0497	0.3079	0.4392	0.0980
CNN-p	*0.2319*	*0.3129*	*0.0708*	0.3347	0.4630	0.1087
CNN	**0.2455***	**0.3313****	**0.0770****	**0.3492****	**0.4744****	**0.1123**

RankSVM with the same features does not do well

Our CNN beats the baseline, significantly... but not by much

Responding to a Request Tomorrow



Microsoft

Team

- Nicola Cancedda
- Yvonne Diep
- Piotr Grudzien
- Ioannis Klapaftis
- Grzegorz Kukla
- Bhaskar Mitra
- Kevin Moynihan
- Silviu Popescu
- Roy Rosemarin
- Christophe Van Gysel (Intern)
- Matteo Venanzi
- Weikun Wang

Questions?

References

[Bendersky and Croft, 2008]	Bendersky, Michael, and W. Bruce Croft. "Discovering key concepts in verbose queries." *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008.
[Cetintas and Si, 2012]	Cetintas, Suleyman, and Luo Si. "Effective query generation and postprocessing strategies for prior art patent search." *Journal of the Association for Information Science and Technology* 63.3 (2012): 512-527.
[Van Gysel et al. 2017]	Van Gysel, Christophe, et al. "Reply with: Proactive recommendation of email attachments." CIKM 2017 and *arXiv preprint arXiv:1710.06061* (2017).
[Xue and Croft, 2009]	Xue, Xiaobing, and W. Bruce Croft. "Automatic query generation for patent search." *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009.
[Xue et al. 2010]	Xue, Xiaobing, Samuel Huston, and W. Bruce Croft. "Improving verbose queries using subset distribution." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.