The MUIR Framework: Cross-Linking MOOC Resources to Enhance Discussion Forums

Ya-Hui An[1,3], Muthu Kumar Chandresekaran[1], Min-Yen Kan[1,2], and Yan Fu[3]

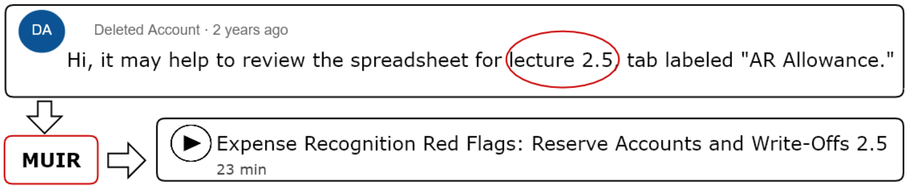
[1] Web IR / NLP Group (WING), National University of Singapore, Singapore
[2] Smart Systems Institute, National University of Singapore, Singapore
[3] Web Sciences Center, School of Computer Science and Engineering, University of Electronic Science and Technology of China, China

Abstract. New learning resources are created and minted in Massive Open Online Courses every week – new videos, quizzes, assessments and discussion threads are deployed and interacted with – in the era of on-demand online learning. However, these resources are often artificially siloed between platforms and artificial web application models. Facilitating the linking between such resources facilitates learning and multimodal understanding, bettering learners' experience. We create a framework for MOOC Uniform Identifier for Resources (MUIR). MUIR enables applications to refer and link to such resources in a cross-platform way, allowing the easy minting of identifiers to MOOC resources, akin to #hashtags. We demonstrate the feasibility of this approach to the automatic identification, linking and resolution – a task known as Wikification – of learning resources mentioned on MOOC discussion forums, from a harvested collection of 100K+ resources. Our Wikification system achieves a high initial rate of 54.6% successful resolutions on key resource mentions found in discussion forums, demonstrating the utility of the MUIR framework. Our analysis on this new problem shows that context is a key factor in determining the correct resolution of such mentions.

Keywords: Digital Library · MOOC · Learning Resource · Unique Resource Identifier · DOI · MUIR.

1 Introduction

Digital libraries for open knowledge goes beyond the scholarly library and extends into the pedagogical one [9]. While participation in Massive Open Online Courses (MOOCs) and online learning has expanded [5, 8, 13, 14], the methods by which learners participate in these classes has still been confined to the limitations of the Learning Management Systems (LMS) [4, 6]. Such LMSes often have separated and distinct views of each form of learning resource – discussion forums, lecture videos, problem sets, homeworks – where cross-linking resources is difficult or impossible to achieve. Learners "cannot see the forest for the trees" when concepts are siloed and easy cross-referencing is impeded.



Deleted Account · 2 years ago

Hi, it may help to review the spreadsheet for lecture 2.5 tab labeled "AR Allowance."

Expense Recognition Red Flags: Reserve Accounts and Write-Offs 2.5

23 min

Fig. 1. Crosslinking a lecture resource mention in a discussion forum.

A concrete instance of this is in the discussion forum, where both instructors and students co-construct arguments to support critical thinking and knowledge [2, 7]. Students often reference a certain quiz, this week's lecture or a particular slide, as in Figure 1. Automatically hyperlinking such mentions to the target resource brushes and links the two endpoints, facilitating the contextualization of course materials across disparate views. To address this, we introduce and reduce to practice a pipeline that adds appropriate hyperlinks to natural language mentions of MOOC resources in discussion forums – a task known as *Wikification*, named after the same task which was first applied to Wikipedia.

In addressing this challenge, we needed to also propose an important standalone contribution: a framework for MOOC Uniform Identifier for Resources, which we name MUIR[4]. The MUIR framework is a two-component framework that pairs a transparent, guessable URL syntax for learning resources with a best-effort resolver that connects MUIR identifiers to their target resource. Best thought of as a hybrid between bibliographic records that identify a scholarly work, and the Digital Object Identifier that gives a resolution, our MUIR framework facilitates the cross-linking functionality that allows for the Wikification of natural language mentions in learner and instructor discourse.

MUIR also facilitates resource discovery. As a central harvester, the MUIR resolver components crawls MOOC platforms for resources and can expose related course material across different providers, formulating a MOOC domain Linked Open Data (LOD) [3], which creates typed links between data from different sources. This helps to address learning resource reuse, a problem that has been exacerbated with exponential success of MOOCs [18]. Without an aggregation service like MUIR, each MOOC LMS platform is siloed: having its own resource identifier schema that is non-portable, opaque and non-interpretable.

We demonstrate the use of the MUIR framework for the application of Wikification. In this case study, our Wikification application recognizes mentions to publicly exposed resources, and generates short form references to those resources which the framework resolves and forwards links.

[4] MUIR refers to **M**OOC **U**niform **I**dentifier for **R**esources as well to the eponymous framework that creates such identifiers.

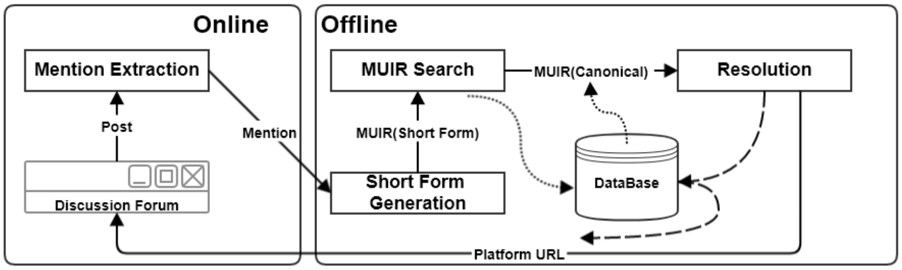


Fig. 2. MUIR System Architecture: (l) online system, (r) offline harvester and resolution components.

2 Related Work

The MUIR framework contributes to both the topics curation and indexing, as well as identification schemes. We review these areas in turn.

Curation and Indexing. Both MOOCs and Learning Resource collection and indexing have prior work. MOOC List[5] curates a commercial, faceted indexing website to find current MOOC offerings. More general and academically inclined, MERLOT[6] achieves broader goals for thousands of learning resources for K–12 and tertiary education, for learners, educators, and faculty development for specific discipline. It acts as both an aggregator of submitted content for peer curation as well as a focal point for gathering the community concerning these resources [11]. MERLOT allocates a unique identifier to each material submitted as a pairing of a unique *'materialId'* and an *'entryType'*. More recently, the OpenAIRE project [1] aggregates metadata about scholarly research – projects, publications, people, organizations, etc.) – into a central information space.

Identifier Schemes. Wikification uses MUIR to cross-link resources, creating a MOOC domain-specific form of Linked Open Data (LOD) [3, 10]. It is a method of publishing to create and publish typed links between data entities from different sources, so that the data can be interconnected and put to better use. The MUIR scheme aims to aggregate resources across platforms and should be persistent, transparent and resolvable for various providers. We are informed of the the design by related resource identifiers such as PURL, DOI, Dublin Core and general bibliographic metadata.

A Persistent Uniform Resource Locator [15] (PURL) provides a single layer of indirection built over the standard URL protocol for web addressing. PURLs solve the problem of transitory URIs through their indirection, but omit any guidelines or enforcement of the identifier minting schema; the choice of identifier is up to the minting agent, somewhat akin to custom URL shorteners such

[5] https://www.mooc-list.com/

[6] or “Multimedia Educational Resource for Learning and Online Teaching”, https://www.merlot.org/

I. MUIR (Short Form, Transparent (sample)): `www.example.org/accounting-analytics/Week 2/lecture/2-5`

II. MUIR (Canonical Transparent): `www.example.org/Coursera/accounting-analytics/1480320000000/Brian J Bushee&Christopher D. Ittner/Videos/expense-recognition-red-flags-reserve-accounts-and-write-offs-2-5`

III. MUIR (Opaque): `www.example.org/id/1239jdn3oni3123s`

IV. Coursera URL: `www.coursera.org/learn/accounting-analytics/lecture/1UzkX/expense-recognition-red-flags-reserve-accounts-and-write-offs-2-5`

Fig. 3. A Coursera learning resource URI in MUIR's threefold identifier scheme.

as `bit.ly` and `tiny.cc`. The Digital Object Identifier [12] (DOI) schema goes further, not bound by any dependent protocols (e.g., HTTP for PURLs) and admits different authorities (e.g., different journal publishers) and distributed and hierarchical resolution via its use of the handle system. Our MUIR proposal is technically a PURL service, where our effort has been to create strong guidelines for the identifier portion of the schema.

Both Dublin Core[17] (DC) and bibliographic metadata are flexible containers that specify preferred (or mandatory) metadata attribute–value fields for different types of materials, such as *title* or *contributor*. Unlike PURL and DOI which are opaque, MUIR opts for transparent identifiers, taking the cue from DC and bibliographic metadata. The components of a MUIR encode the metadata values directly as part of the URL syntax for the identifier, and uniformly across various LMS providers.

3 The MUIR Framework

"When we try to pick out anything by itself, we find it hitched to everything else in the universe" — John Muir

Our uniform identifier scheme for MOOC learning resources, embodies the American naturalist John Muir's insight that everything is interconnected. In creating MUIR, our aim is to objectify MOOC resources so that they can be inventoried, referenced and subsequently better "hitched" to other resources, in the spirit of LOD, creating a densely tangled web of knowledge crucial for the contextualization of learning. We discuss the desiderata for our MUIR schema, while relating it to the practices of related work.

We motivate this section by working through the elements of a hypothetical MUIR associated with a learning resource from Coursera representing a specific lecture on accounting analytics:

1. Indirection. MUIRs provide two layers of indirection over actual resolvable resources such as a Coursera discussion forum, or a quiz hosted on a course on EdX. The first layer serves as a semantically transparent, short form where fields can be omitted and the search functionality of MUIR invoked to form the best-effort resolution to the canonical form. Similar to the simplicity of #hashtags, the MUIR short form encourages direct use by humans, later to be resolved to a canonical form or directly to the platform URL via best-guess relevance search.

The second layer of indirection (from the canonical form to the platform URL) provides both a uniform access mechanism to the resources that is platform- / provider-independent. As with PURLs, it also lends itself to preservation, having a single authority for resolution. Both the canonical form and the opaque form map one to one to the platform instance.

2. Transparent. Unlike traditional schema that use succinct opaque identifiers to serialize and identify objects, MUIR takes the cue from bibliographic systems that admit multiple, value–attribute fields to name resources. Much like how Dublin Core mandates certain fields be specified, MUIR also splits fields into required (*Resource Title*, *Resource Type*, *Course Name*, *Session Date*, *Instructor(s)*, *Institution*, *Source Platform*) and optional categories (*Other Elements*). The short form MUIR invokes search by the resolution system to find the most appropriate learning resource, akin to search in a web search engine or an online public access catalog.

3. Comprehensive. MUIR's resource type categorizes the most common learning resources exposed in MOOCs. We survey learning resources provided on 29 worldwide MOOC platforms to inventory the common learning resources exposed, and map these forms to MUIR's *Resource Type* (Table 1). *Videos* present the lecture content. *Slides* provide the lecture content for download and separate review, often aligned to those in the video. *Transcripts* of the videos are sometimes available for various languages, often for other languages than the one used in the video. *Assessments* capture any form of assessments, exercises, homeworks and assignments that aim to self-diagnose the learners' knowledge commitment of the course content. *Exams* evaluate the knowledge and/or skills of students, including quizzes, tests, mid-exams and final examinations. *Readings* optionally provide a list of other learning resources provided by courses. *Additional Resources* help to catch other materials made available for specialized discipline-specific courses. For example, computer programming courses can provide program files for reference.

4. Stable. In addition to standard descriptor-like identifier structure, MUIR also has an alternate serial identifier syntax that is opaque and succinct, permitting short references that are permanent, as in the final MUIR opaque identifier in Figure 3. Thus there can be many MUIR short form, transparent descriptors that map to a single unique opaque identifier.

3.1 Collected Dataset

We operationalize our MUIR framework by creating a series of crawlers to proactively collect learning resources from MOOC platforms. In the remainder of the paper, we study using MUIR against a subset of crawled resources from Coursera as a proof of concept. Our Coursera corpus, collected at January 31, 2017,

No. Platform	Country	Scale (C /L)	V.	S.	E.	Q.	Tr.	HW.	Asg.	Ass.	Ex.	Re.	Art.	Pro.	Add.
1. Coursera	US	2000+ /25M+	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
2. edX	US	950+ /14M+	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓
3. Udacity	US	200+ /4M+	✓			✓	✓								
4. FutureLearn	UK	400+ /6.5M+	✓			✓			✓		✓		✓		✓
5. iversity	GER	50+ /0.75M+	✓							✓					
6. Open2Study	AU	45+ /1.1M+	✓	✓		✓	✓			✓					
7. Acumen+	US	34+ /0.3M	✓				✓		✓			✓			✓
8. P2PU	US	200+ /—	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
9. Academic Earth	US	600+ /5.8M+	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10. Alison	IE	1000+ /11M+	✓												
11. Athlete Learning Gateway	CH	27+ /14K+	✓				✓					✓			
12. Canvas Network	US	200+ /0.2M+	✓			✓	✓		✓						
13. Course Sites	US	493+ /—	✓			✓			✓			✓			✓
14. KhanAcademy	US	— /57M+	✓				✓	✓		✓	✓				
15. Open Learning	JP	30+ /—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
16. OpenupEd	EU	190+ /—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
17. Saylor	US	100+ /—	✓		✓	✓					✓		✓	✓	
18. Udemy	US	— /20M+	✓			✓									✓
19. CNMOOC	CN	600+ /—	✓			✓									
20. Complexity Explorer	US	11+ /—	✓			✓	✓	✓							✓
21. Ewant	TW	600+ /20K+	✓	✓		✓									
22. Janux	US	20+ /31K+	✓			✓	✓		✓						✓
23. Microsoft Virtual Academy	US	800+ /—	✓	✓		✓	✓								
24. NTHU MOOCs	TW	46 /—	✓			✓			✓						
25. Stanford Online	US	100+ /—	✓	✓	✓	✓		✓	✓	✓					✓
26. XuetangX	CN	1300+ /9M+	✓	✓	✓	✓		✓	✓	✓					✓
27. icourse163	CN	1000+ /—	✓		✓	✓		✓	✓						
28. FUN	FR	330+ /1M+	✓			✓					✓				
29. FX Academy	ZA	10+ /—	✓			✓						✓			✓
# of platforms w/ Resource Type			29	11	10	24	15	11	16	11	9	8	5	6	15
Mapping to MUIR's Resource Type			V.	S.	E.		T.	Ass.				Re.		Add.	

Table 1. Prevalence of resource types exposed on global MOOC platforms. 'Scale' indicates # of courses / # of learners. The subsequent columns on top represents videos, slides, exams, quizzes, transcript, homeworks, assignments, assessments, exercises, readings, articles, programming scripts and additional materials, respectively. Each resource type is mapped to one of MUIR's canonical resource types (bottom row).

includes all posts and resources of 142 courses that had already completed, totalling 102,661 posts and 11,484 learning resources spanning all 7 resource types.

4 Discussion Forum Wikification

We operationalise the MUIR framework through the task of *discussion forum Wikification*. Our system for forum Wikification extracts and hyperlinks mentions of learning resources in student posts as shown in Figure 1.

The skeptic might ask: Is Wikification meeting a real demand for crosslinking learning resources? To answer this, we wish to calculate the number of mentions that are actually present in discussion forum. Let us assume that mentions to the seven resource types do contain a descriptive keyword. While the presence of these keywords may not necessarily denote an actual mention (i.e., "*I have a question*"), the percentage of posts that contain the relevant keywords serves as an upper bound for the number of mentions. Restricting our examination to content subforums (excluding forums for socializing; e.g. '*Meet & Greet*' and '*General Discussion*'), we find that approximately 15,529/69,025 = 22.5% posts contain one or more keywords. Restating, about 1 of 4 posts in discussion forums potentially have mentions that need Wikification. So there is a real need that we address with Wikification.

The process presumes that the MUIR system has proactively crawled and indexed MOOC resources, as previously discussed. We reduce the problem into 4 concrete phases as shown in Figure 2: 1) Mention Extraction: mention identification, 2) Short Form Generation: MUIR short form construction, 3) MUIR Search: MUIR short form to canonical form resolution, 4) Resolution: forwarding the request to the platform URL. Note that the first two phases take place outside of the MUIR framework, in our Wikification application that processes discussion forums. We step through these four phases in turn to illustrate how the MUIR framework interacts with the Wikification process.

Phase 1: Mention Extraction. Wikification begins by identifying important mentions from a post of a course. As natural language mentions can occur in an infinite variety, in this initial study, we constrain the problem scope to identifying only **S**ingle, **C**oncrete, w**I**thin-course entities (or SCI). As counterexamples, references to collective entities (i.e., "the quizzes"), specific topics taught within a course (similar to keywords, i.e., "corporate risk") fall outside the scope of our SCI definition.

Analyzing actual SCI mentions in discussion forums, such as *"lecture 2.5"* in Figure 1 and those in Figure 4 show us that SCI entities do lend themselves to be captured by a simple regular expression matching with a keyword followed by a numeric offset. We thus programmatically find and delimit such mentions as spans for hyperlinking. This solution, although overly simplistic, serves well as a starting point for Wikification. We revisit this decision later in our evaluation.

Phase 2: Short Form Generation. For each mention, Wikification generates a MUIR short form programmatically. The short form is used to split the mention into component words, using which our algorithm maps them to fields in the MUIR short form. Inferrable missing components are added by the context of the hosted discussion forum. Continuing with our running example, this stage takes the mention *"lecture 2.5"* that appears in an *Accounting Analytics* course on Coursera, and constructs the short form I in Figure 3, where the mention's text of {*"lecture"*, *"2"*, *"."* and *"5"*} constructs the s_4 and s_5 short form components: the relative block number (2−5 denotes module 2 lecture 5), and remaining components (s_2 and s_3) are inferred from context:

$$\underbrace{\texttt{www.example.org}}_{s_1}/\underbrace{\texttt{accounting}-\texttt{analytics}}_{s_2}/\underbrace{\texttt{Week2}}_{s_3}/\underbrace{\texttt{lecture}}_{s_4}/\underbrace{\texttt{2}-\texttt{5}}_{s_5}$$

Here, s_1 is the MUIR resolver host, s_2 is the course name, s_3 is the forum name (usually the week number) of the post, s_4 is the resource type and s_5 represents the relative block number.

Phase 3: MUIR Search. A click on a short form requests the resource from MUIR resolver. This search process is the first layer of indirection, combining the post information in the MUIR database from which MUIR obtains additional peripheral information (platform, session date and instructor(s) name) about the post that embeds the mention. The search process first utilizes the origin post data {source platform, s_2, session date and instructor(s) name} to locate the hosting course's context. The remainder of the short form (s_4 and s_5) are used to match the resource type and name in a full text search, where exact matches are favored. The resolver searches its index of canonical MUIRs using this custom search logic to match with the short form and deems the best match its resolution. As in the running example, this process matches the MUIR short form I to the MUIR canonical form II:

$$\underbrace{\texttt{www.example.org}}_{f_1} / \underbrace{\texttt{Coursera}}_{f_2} / \underbrace{\texttt{accounting} - \texttt{analytics}}_{f_3} / \underbrace{\texttt{1480320000000}}_{f_4} / \underbrace{\texttt{BrianJBushee\&Christopher}}_{f_5}$$

$$\underbrace{\texttt{D.Ittner}}_{f_5} / \underbrace{\texttt{Videos}}_{f_6} / \underbrace{\texttt{expense}-\texttt{recognition}-\texttt{redflags}-\texttt{reserve}-\texttt{accounts}-\texttt{and}-\texttt{write}-\texttt{offs}-\texttt{2}-\texttt{5}}_{f_7}$$

Here, f_1, f_3 and f_6 are migrated from the short form, and the remaining fields have been imputed from context: f_2, f_4, f_5 and f_7 give the source platform, the session date, instructors' names, and the slug name of the resource, respectively.

Phase 4: Resolution. This final phase is simple, as the canonical MUIR maps one-to-one with a platform URL, through a hash table lookup. This process maps the running example's canonical form II to the platform-specific URL IV through the second layer of indirection.

5 Wikification Evaluation

We believe the MUIR identifier framework is useful on its own right, but it is hard to evaluate its intrinsic utility. We instead evaluate extrinsically, assessing the utility of MUIR as a component within discussion forum Wikification. Specifically we ask ourselves the following research questions (RQ):

RQ1. What is the coverage rate for posts that actually contains mentions?
RQ2. How accurate is the resolution for different resource types?

RQ1: Mention Coverage. With a full annotation of the dataset we could conclusively measure the coverage of our regular expressions in capturing actual natural language mentions to SCI. However, the effort for full annotation is infeasible, and instead we randomly sample ∼1,000 posts to check the actual coverage of our Wikifier syntax. We note that it can be unintuitive for annotators to identify whether a word, phrase or sentence is a mention, so we employed two independent annotators to reduce bias. Results for this sample annotation are shown in Table 2.

Annotator ID	# of Posts	# of posts identified as having mentions	# Extracted by our Wikifier	# Correct	Coverage
Annotator 1	1,087	156	5	5	14.4%
Annotator 2	1,087	175	5	5	16.1%
Overall	1,087	196 (Union)	5	5	18.0%

Table 2. Mention extraction coverage.

YES: $\langle m_1 \rangle$ Is it just me or were some questions on Quiz 2 a surprise? There were a few questions that were not discussed in the lesson plan.
YES: $\langle m_2 \rangle$ Hello, I just would like to note that on 12:30 in the answer to question 3 in the lecture 2.4 it says that the network is deadlock-free, whereas ...
NO: $\langle m_3 \rangle$ The last item, that is "Probability Models for Customer-Base Analysis.pdf", in the Resources &gt; Additional Readings by Week section for Week 3 is not accessible.
NO: $\langle m_4 \rangle$ I'm working on the programming assignment for ML, week 2. I successfully submitted answers to the obligatory questions.
NO: $\langle m_5 \rangle$ At around 5:00 in the lecture, we see that the regularization term in the cost function is summed from 1 to L-1. Shouldn't this be 2 to L?
NO: $\langle m_6 \rangle$ Hello. I wanted to use "e" as a number for ex.2/week3. It didn't work, and I didn't find useful help with "help exponent".

Fig. 4. Actual resource mentions in our 1,087 sample sized dataset, illustrating the variety of expressions. Our Wikification currently handles the first two mentions.

In our 1K sample of posts, 18% of posts or more contain mentions to learning materials. This is significant, as it shows that there is much potential to better interlink resources, even just for the silo of discussion forums. In these sampled posts, our Wikifier matched 5 mentions, which were all actual mentions (correct). This result shows that our <"*keyword*" + number> pattern has high precision but suffers from low recall, covering only about 2.6% of possible mentions.

How can we improve mention extraction coverage? We examine the causes for the coverage disparity, where the parenthetical percentage is determined over the same sampled data.

1. **Implicit Contextual Knowledge ($\sim$45% of errors).** In sequential posts, posters often refer to the content from the previous posters, and refer using demonstrative pronouns such as '*this*', '*that*' or '*the*'. Without context knowledge, our prototype simply does not capture such mentions, such as in '*that video you mentioned*'.
2. **Named Reference ($\sim$30% of errors).** Direct use of the resource name – especially for videos, slides and quizzes – makes such mentions impossible to capture, without predicating prior MUIR lookup (cf m_3 in Figure 4 or '*the problem "Hashing with chains"*').

Resource	# of instances	P_I	P_II
Videos	89	71.9%	57.3%
Slides	27	74.1%	33.3%
Exams	718	83.0%	53.3%
Assessments	12	50.0%	25.0%
Total	846	81.1%	54.6%

Table 3. Resolution Accuracy Evaluation. Only mentions to 4 MUIR types are present in our Coursera subset. P_I represents precision of Annotation I and P_II is for Annotation II.

3. **Informal Expressions (∼15% of errors).** Colloquial expressions abound (Figure 4's m_4 and m_6) and fall outside the current scheme. Adding regular expressions to capture these would improve coverage at the cost of precision.

RQ2. MUIR Resolution Accuracy. The other component that needs evaluation is Phase 3, MUIR Search. Given the short forms that are generated by Wikification, MUIR Search connects the short form to a (hopefully correct) platform URL.

We offer two evaluations that give complementary data on the resolution accuracy, shown in Table 3. Comparing P_I against P_II, the accuracy of Annotation I is generally better than Annotation II. That is because Annotation I is generated only by depending on the information of mentions and the limited relevant information of posts, foregoing the implicit contextual knowledge of the previous and subsequent posts. This gives an upper-bound for how well mentions are actually resolved by our simple search logic. But in Annotation II when we annotate the ground truth test data, we consider all of the context of the mentions including the content around the mentions and other posts in the same thread. This is a realistic evaluation on the full complexity of the problem.

The results are best analyzed jointly. We see that the mentions we capture are easy to extract (higher performance on Annotation I), but hard to resolve without context (lower performance on Annotation II). The accuracies for four *Resource Types* have different degrees of reduction. But the results are encouraging: our prototype, even with its simple logic, can already handle almost 55% of learning resources.

As we did for RQ1, we further categorized a rough cause to the errors in the resolution process:

1. **Mentions needing context to resolve against multiple matches (∼ 20% of errors):** Learners may write mentions such as "*lecture 4.5*", where "*4*" and "*5*" are used by MUIR Search but could refer to different lectures that both have textual components "4" and "5" in their slug name.
2. **Multiple potential targets (∼ 70% of errors):** Even considering context, certain mentions are still ambiguous. If a mention states "*question 3*" but there are multiple quizzes within the context, all which have a Question 3, the target is ambiguous. MUIR can only guess in this case.

3. **Errors in mention extraction ($\sim$ 10% of errors):** These are cascaded from the Phase 1 process of mention extraction. Examples include *partial mention extraction* ("*lecture's 2 transcript*" may be written by a learner, but only "lecture 2" was detected) and *informal reference* (*cf* m_6 in Figure 4).

Both RQ1 and RQ2 discussions clearly point forward in the direction of improving coverage, especially in Phase 1, as such errors cascade. A clear direction is to incorporate contextual knowledge: our current work thus aims to incorporate such knowledge by the machine reading of the posts, by leveraging recurrent neural network based learning models [16] currently making much impact in natural language processing research. This will help the Wikification process by both capturing more natural mention expressions and minting better Phase II MUIR short forms that better facilitate correct resolution downstream.

We note that mention extraction can also be facilitated by introducing linking conventions, similar to #hashtags. MUIR's short form can be further facilitated by the future learner's explicit triggering when writing their posts: i.e., "*I have a question about #video5*", where mention identification are solved by the learner.

6 Conclusion

For a learner to see the forest for the trees requires seamless interlinking of learning resources. Discussion forum Wikification takes us closer towards this goal. Our prototype shows the feasibility of the approach for simple mention types, and further motivates research on better mention identification and search resolution of such mentions.

Underlying this development is our core contribution of the MUIR framework for identifying and referencing the burgeoning set of MOOC resources being generated by the community. Our solution hybridizes best practices among ease-of-use descriptions, search practices and the persistence and identification standards. Our work aims to catalyse work towards making linked open data a closer reality for the world's learners.

Acknowledgement

This research is funded in part by NUS Learning Innovation Fund – Technology grant #C-252-000-123-001, and also in part by the scholarship from China Scholarship Council (CSC), National Natural Science Foundation of China under Grant Nos.: 61673085 and 61433014, and UESTC Fundamental Research Funds for the Central Universities under Grant No.: ZYGX2016J196. We also thank the anonymous reviewers for their useful comments.

References

1. Ameri, S., Vahdati, S., Lange, C.: Exploiting interlinked research metadata. In: International Conference on Theory and Practice of Digital Libraries. pp. 3–14. Springer (2017)

2. Andresen, M.A.: Asynchronous discussion forums: success factors, outcomes, assessments, and limitations. Journal of Educational Technology & Society **12**(1), 249 (2009)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data-the story so far. International journal on semantic web and information systems **5**(3), 1–22 (2009)
4. Dalsgaard, C.: Social software: E-learning beyond learning management systems. European Journal of Open, Distance and E-Learning **9**(2) (2006)
5. Hew, K.F., Cheung, W.S.: Students and instructors use of massive open online courses (moocs): Motivations and challenges. Educational research review **12**, 45–58 (2014)
6. Mahnegar, F.: Learning management system. International Journal of Business and Social Science **3**(12) (2012)
7. Marra, R.M., Moore, J.L., Klimczak, A.K.: Content analysis of online discussion forums: A comparative analysis of protocols. Educational Technology Research and Development **52**(2), 23 (2004)
8. Martin, F.G.: Will massive open online courses change how we teach? Communications of the ACM **55**(8), 26–28 (2012)
9. McAuley, A., Stewart, B., Siemens, G., Cormier, D.: The mooc model for digital practice (2010)
10. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: Dbpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. pp. 1–8. ACM (2011)
11. Moncada, S.M.: Rediscovering merlot: a resource sharing cooperative for accounting education. Journal of Higher Education Theory and Practice **15**(6), 85 (2015)
12. Paskin, N.: Digital object identifier (doi®) system. Encyclopedia of library and information sciences **3**, 1586–1592 (2010)
13. Peña-López, I., et al.: Giving knowledge for free: The emergence of open educational resources (2007)
14. Seely Brown, J., Adler, R.: Open education, the long tail, and learning 2.0. Educause review **43**(1), 16–20 (2008)
15. Shafer, K., Weibel, S., Jul, E., Fausey, J.: Introduction to persistent uniform resource locators. INET96 (1996)
16. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
17. Weibel, S., Kunze, J., Lagoze, C., Wolf, M.: Dublin core metadata for resource discovery. Tech. rep. (1998)
18. Zemsky, R.: With a mooc mooc here and a mooc mooc there, here a mooc, there a mooc, everywhere a mooc mooc. The Journal of General Education **63**(4), 237–243 (2014)