Team Tutoring Automated Content Discovery and Learning Objective Alignment

Benjamin Bell[1], Keith Brawner[2], Elaine Kelsey[1], & Debbie Brown[1]

[1] Eduworks Corporation, Corvallis OR 97333, USA
[2] U.S. Army Combat Capability Development Command – Soldier Center – Simulation and Training Technology Center
benjamin.bell@eduworks.com

Abstract. Despite advances in team tutoring, creating and maintaining these systems remains costly and time-consuming. Authoring tools accelerate learning systems development by supporting instructional design tasks like sequencing, feedback, adaptation, and assessment. To scale learning to meet team training needs, however, tutoring systems must be capable of incorporating content that is broad and readily adaptable as learning needs shift in response to equipment upgrades, changes in tactics, evolving threats, and operations in new theaters. An additional and significant challenge is the emerging emphasis on team training. Authors of team tutoring face complicated content management tasks related to distinguishing content that supports individual skills and content aligned with team skills, as well as identifying content associated with specific roles within a team. To help developers of team training more efficiently find and maintain relevant content, automation is needed that supports the analysis of information and its alignment with team and individual learning objectives (LOs). In this paper we introduce Machine-Assisted Generation of Instructional Content (MAGIC), a new authoring aid to help training developers find, organize, and curate resources aligned with desired LOs. MAGIC analyzes source documents and extracts content that aligns with specified learning objectives, lending much-needed support for team training development by distinguishing between individual and team LOs. Moreover, MAGIC identifies content associated with specific roles within a team. We present promising early findings and discuss work in progress extending MAGIC to analyzing task lists, game scenarios and orders.

Keywords: Team Tutoring, Authoring Tools, Machine Learning.

1 Introduction

The process of authoring Intelligent Tutoring Systems (ITSs) is labor-intensive and requires highly-specialized skills. Authoring environments, such as the Generalized Intelligent Framework for Tutoring (GIFT) [15] can accelerate ITS development by supporting instructional design tasks like sequencing, feedback, adaptation, and assessment. With growing demand for team tutoring in support of rapidly-evolving Army requirements, tutors must be able to scale learning to meet team training needs, be capable of incorporating broad content; and offer instructional value for both individual Soldiers and teams [4, 11, 13, 14, 16, 17] (. A key need is to help ITS authors efficiently find and maintain relevant content, and to assist authors with discriminating between content supporting individual learning objectives and team learning objectives. Addressing this need efficiently calls for automation that supports the analysis of information and its alignment with learning objectives (LOs) [2].

In this paper we introduce a new authoring aid, designed to be incorporated within GIFT, to help ITS developers find, organize, and curate resources aligned with desired individual and team learning objectives. Machine-Assisted Generation of Instructional Content (MAGIC) analyzes source documents and extracts content that aligns with specified learning objectives. MAGIC additionally lends much-needed support for team training development by performing this alignment for both individual and team learning objectives. Building on and extending existing artificial intelligence (AI) and natural language processing (NLP) techniques, MAGIC streamlines content alignment, distinguishes between individual and team content, and helps extend the reach of intelligent tutors to meet Army team training demands.

2 Team Tutoring: Scaling Challenges

Scaling virtual training for teams to fully address Army needs requires tools and techniques for efficiently creating team tutoring simulations. While authoring tools such as GIFT support numerous instructional design tasks (organizing content into modules, sequencing content, creating tailored feedback), finding and organizing content that aligns with desired learning objectives remains a labor-intensive process that takes place outside of the authoring process. Achieving scale means that virtual training must span broad content. In addition to scaling, content must maintain relevance in order for virtual training to be readily adaptable as learning needs shift in response to equipment upgrades, changes in tactics, evolving threats, and operations in new theaters.

To create team training, authors must navigate complicated content management tasks related to distinguishing content aligned with both *individual* skills and *team* skills, as well as trying to identify content associated with specific roles within a team. Team members must attend to their role, function as a member of the group, and ensure group success. Creating and maintaining virtual team training systems thus remains costly and time-consuming. To help developers of team training find and tag relevant content more efficiently, automation is needed that supports analysis of content and its alignment with team and individual learning objectives.

The content alignment approach in MAGIC answers this need by helping training developers find, organize, and curate resources aligned with desired learning objectives. MAGIC analyzes source documents and extracts excerpts of content that aligns with specified learning objectives, and performs this alignment for both individual and team learning objectives. Moreover, MAGIC identifies content associated with specific roles within a team. A schematic depiction of MAGIC is shown in Figure 1.

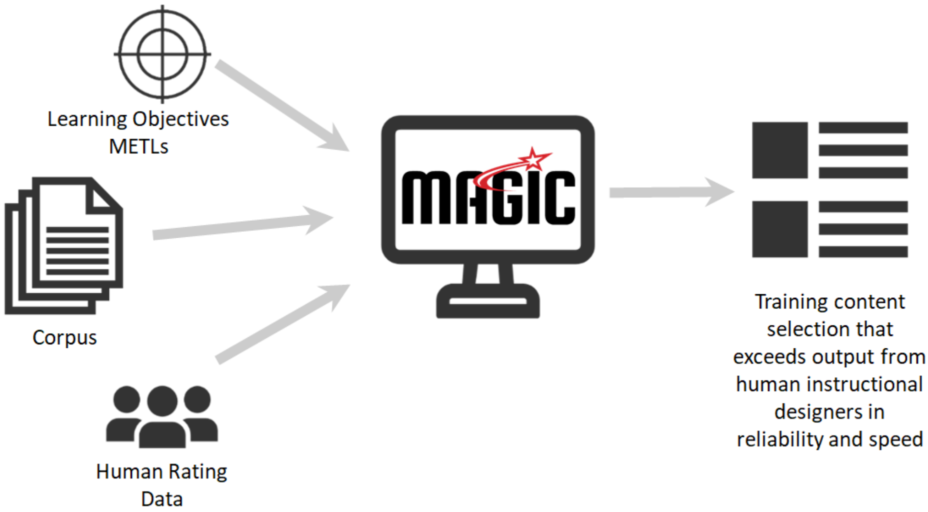

Fig. 1. MAGIC at a glance.

3 Supporting Content Alignment

3.1 Text Analysis Approach and Challenges

A challenge MAGIC addresses is matching content excerpts to a learning objective (typically a short text string) rather than to a topic (typically supported by larger amounts of descriptive text). To address this difficulty, we extended existing work in word embedding approaches (e.g. Word2Vec, GLoVe) [9, 10], to develop a new technique we refer to as concept embedding. The approach first involves parsing an input corpus of documents to detect entities and relations as short phrases (rather than as individual words) using TensorFlow (SyntaxNet-style) for dependency parsing along with traditional ontological approaches [5].). In the next step, we build corpus models using the resulting dependency trees as the input into distinct entity and relation embedding models, where ‘concepts’ are defined as tight clusters of phrases in the resulting vector spaces [7]. By mapping entities and relations separately, and then linking them through a combined (modified W2V-SG) model, we are able to instantiate concepts as tight clusters of phrases that exist in the resulting entity and relation vector spaces. For example, this approach might instantiate the concept “Santa Claus” as associated with “Jolly Old St. Nick” and “the fat man in the red suit” [8, 12].

This concept embedding approach gives MAGIC the ability to extract a richer description of meaning from very short text strings (namely, learning objectives). In our use case, the approach is applied in multiple steps to perform excerpt extraction:

- Extract entities and relations from the LOs
- Generate an embedding space

- Map entities to concepts
- Use any available context to disambiguate between concepts
- Map documents to the concept space (both concept and topic levels)
- Match concepts in each LO to concepts in the corpus
- Rank results based on match to both entity-concepts & relation-concepts by LO

To discriminate between individual or team LO types, we apply a hybrid Machine Learning (ML) approach combined with syntactic-semantic patterns [6]. For the ML component, we extract the semantic and syntactic features and test using Naïve Bayes and Support Vector Machine (SVM) classification techniques, which produces similar results. However, these two approaches are more accurate and require less training data than either a Convolutional Neural Network (CNN) or Recurrent Neural Network (RNN) implementation. For the Syntactic-Semantic component, we extract combined syntactic-semantic features using SyntaxNet with TensorFlow, matched via the pattern library. We achieved the best results by applying both the ML and Syntactic-Semantic Pattern approaches and then using context-specific heuristics (derived from features of the source document and larger source text) to resolve any disagreements.

To identify an appropriate team role for an excerpt, we determined that the link to LOs/competency frameworks provides important role implications as well as a predefined list of possible roles. Our approach was to expand each role into a Concept using the Concept Embedding Model, and then to apply a similar matching approach. We continue to take steps to improve results with role assignment by using human-labelled data to detect discourse and semantic-syntactic markers for a list of common domain-specific roles. The application of a supervised learning layer trained with human-tagged samples and an ontological model of military domain roles is expected to further enhance MAGIC outcomes, with a goal of achieving accurate extractions and tag selections more often than the human raters.

3.2 Machine Learning Models

MAGIC uses ML and NLP techniques to train algorithms that associate content with learning objectives, tag content as having individual or team relevance, and associate content with specific team roles when applicable. We developed three sets of ML models for our initial research and testing: (1) unsupervised general models trained using Wikipedia and the New York Times Annotated Corpus to map concepts; (2) unsupervised domain-specific models trained with military-sourced documents to define domain-specific concepts; (3) supervised, domain-specific models trained with human-tagged data from a team of instructional designers and subject-matter experts to enhance outcomes. In the case of the battle drill use cases, we manually created learning LOs outlined as hierarchical task procedures, based on original document text, and manually tagged content with task type and role as depicted in Figure 2. The manually tagged LOs were used to train the ML algorithms for task type and role detection.

COMPETENCY ID	PARENT ID	Standard or Competency	Team or Individual	Role(s)
BD10-S1		*All Soldiers don their protective mask within nine seconds (or fifteen seconds for masks with a hood).*	*Individual*	*All soldiers*
BD10-S1-T1	BD10-S1	Element dons their protective mask.	Individual	All soldiers
BD10-S1-T2	BD10-S1	Element gives vocal or nonvocal alarm.	Team	Designated soldier(s)
BD10-S1-T3	BD10-S1	Element uses the appropriate skin decontamination kit within one minute for individual decontamination, as necessary.	Team	All soldiers
BD10-S2		*Soldiers assume MOPP4 within eight minutes.*	*Individual*	*All soldiers*

Fig. 2. Example learning objectives for a battle drill.

To create the tagged data we used a team of three human raters with instructional design, research, and military backgrounds, led by an expert in instructional design. Raters were trained on the rating task, which included scoring relevance of sections of content to a learning objective and tagging with individual/team and team role identifiers. The resulting tagged data set consists of 3,132 tagged items and was segmented into two corpora: one for training the supervised learning models, and one for evaluating performance of all three ML model sets. The average interrater reliability (n=3) was 81.6% for text selection and extraction, 87.8% for distinguishing team and individual content, and 78% for identifying team roles.

4 MAGIC Proof-of-Concept

4.1 Prototype Design

The MAGIC prototype is designed to illustrate support for a training developer. Using the MAGIC interface, a training developer provides a list of learning objectives and selects the target library (or corpus of documents) to be analyzed as shown in Figure 3. For our initial demonstration of the MAGIC algorithms, we drew learning objectives from battle drills in the Maneuver domain; for the library we used the Central Army Registry (CAR) [1] and the Milgaming portal [3] Training Support Packages (TSPs) to create a collection of over 1,200 documents.

MAGIC then generates a collection of text excerpts from across the selected documents, each tagged by the learning objectives, individual or team types, and team roles the excerpt aligns with. In the current demonstration interface, these results may be viewed, filtered, and compared with human rater results when available (Figure 4). In future work, the toolset will offer more flexible export packaging options designed to integrate into GIFT repository search and authoring components using the MAGIC API.

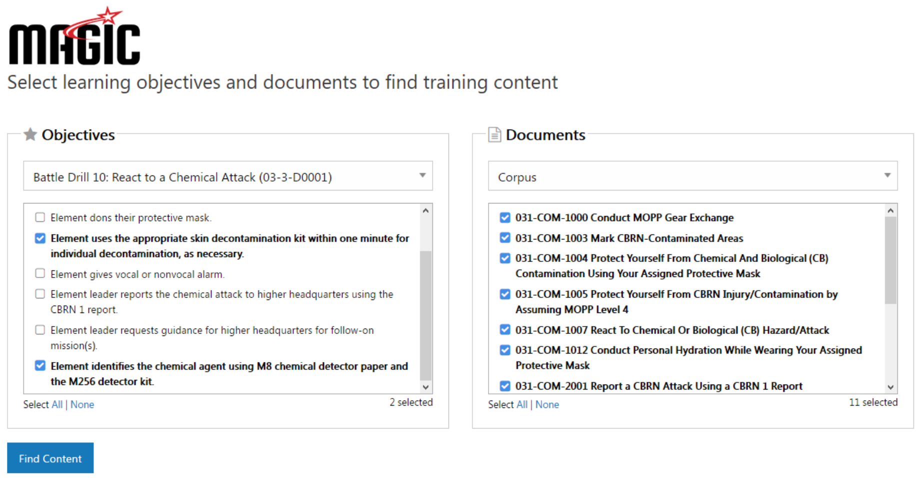

Fig. 3. Selecting learning objectives and corpus documents to configure a content analysis.

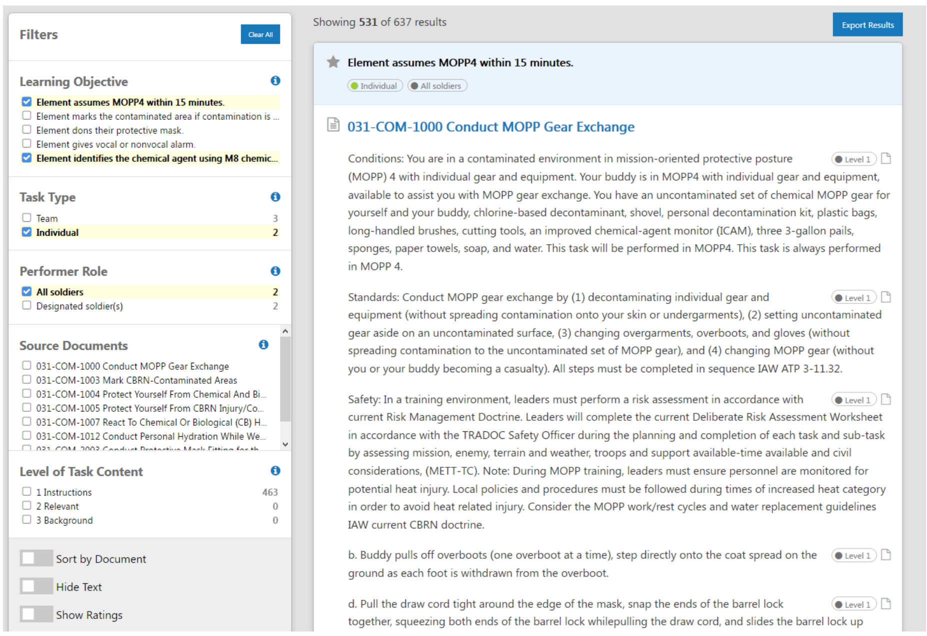

Fig. 4. Filtering content by LO, task type, and role

4.2 Preliminary Results

To provide early metrics of MAGIC's performance, we used the second set of labeled data as a test set. Both the training and test sets comprised approximately 5,000 comparisons of a text excerpt to a LO, and each task was completed by the three independent raters. Interrater reliability was 81.6%.

	Unsupervised Domain-General Model	Unsupervised Domain-General Model + Unsupervised Domain-Specific Model	Unsupervised Domain-General Model + Unsupervised Domain-Specific Model + Supervised Domain-Specific Model
Match 1: Machine agrees with raters when all humans agree	0.827	0.923	0.986
Match 2: Machine agrees with the majority when humans disagree	0.514	0.723	0.769
Match Total	0.762	0.871	0.948

Fig. 5. Preliminary results for each of MAGIC's Machine Learning models.

The results (Figure 5) demonstrate the algorithms performing slightly below human performance when using only the domain-general unsupervised model, at or near human performance when adding the unsupervised domain-specific model, and slightly above human performance with the supervised domain-specific model added.

5 Related Applications

The value offered by MAGIC is the considerable time savings associated with identifying content across a large repository of material that is aligned with specific learning objectives of interest. MAGIC overcomes the algorithmic challenge of matching content to a small sample of text (a LO) through sophisticated concept mapping and ML, and its ability to distinguish individual and team learning content as well as tagging content with specific team roles offers a novel and powerful tool for accelerating the creation of team training. The same infrastructure may be leveraged by recommenders for providing relevant training to Soldiers and teams before or during task performance.

This general framework, however, can be leveraged for discovering content that aligns with constructs other than learning objectives. MAGIC, fundamentally, finds content across a collection of documents that aligns in some way with a desired list of items of interest (and works even with items expressed as short text strings). This capability generalizes to other mission-critical Army functions, two of which we are exploring and referred to previously in this paper.

The first is mission essential task lists (METLs), used across the service branches to distill the number of tasks an organization must train. Commanders analyze the tasks set forth in external directives and only select for training those tasks essential to accomplish their organization's wartime mission. MAGIC offers a workflow for rapidly assembling content that supports an organization’s METL, at any echelon.

The second related application is supporting operational orders (OPORDs) and fragmentary orders (FRAGOs). An OPORD is a plan, following a standard five-paragraph format, that helps subordinate units with the conduct of military operations by describing the situation the unit faces, the mission of the unit, and what supporting activities the unit will conduct in order to achieve their commander's desired end state. A FRAGO is an abbreviated form of an OPORD, issued on a more frequent (often daily) basis that modifies or updates the OPORD without the need to restate the information in the OPORD. We are currently investigating the viability of using MAGIC to comb through document repositories and extract content relevant to a given FRAGO, to support rapid production of briefings and mission plans. This application can be especially important when there have been changes to a body of governing documents that provide the framework within which a FRAGO is to be conducted.

6 Conclusion

With preliminary results already meeting human-rater levels of reliability using the combined unsupervised general and domain specific models, and with the addition of a supervised domain-specific model performing better than the human raters, the MAGIC approach is showing promising results and a path for continued enhancement. Based on these early findings, we see the potential for automated content discovery using LO auto-alignment and text extraction will result in faster, scalable team training development processes. Integration of MAGIC services into the GIFT authoring workflows will propel reuse of training materials, while helping training developers overcome the challenges of distinguishing content supporting team or individual learning and aligning content with specific team roles.

Our next steps in the MAGIC project will include creating a supervised domain-specific model for assigning team roles; incorporating non-text content (such as metadata or automated transcriptions); designing a MAGIC services API; testing and evaluation of MAGIC with authors of team training simulations; and the integration of MAGIC services with Army-selected authoring/CMS/LMS tools. Future work will explore extending the application of MAGIC to related use cases, including training guided by METLs and streamlining the production of briefings and mission plans.

7 References

1. Army Training Support Center (n.d.). Central Army Registry (CAR). https://rdl.train.army.mil/catalog/dashboard. Department of the Army, HQ Training and Doctrine Command (TRADOC) Army Training Support Center.

2. Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., Holub, J. The challenges of building intelligent tutoring systems for teams. In *Proc. of human factors and ergonomics society*, **60**(1), pp. 1981-1985. (2016).
3. Combined Arms Center (n.d.). Milgaming, https://milgaming.army.mil. U.S. Army Com bined Arms Center-Training (CAC-T).
4. Fletcher, J. D., Sottilare, R. A. (2018). Shared mental models in support of adaptive instruction of collective tasks using GIFT. International Journal of Artificial Intelligence in Education, **28** (2) pp.265-285. (2018).
5. Goldberg, Y., Levy, O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722. (2014).
6. Kelsey, E. Goetschalckx, R. Robson, E., Ray, F., Robson, R. Automated tools for question generation. United States Patent Application 20180260472
7. Levy, O., Goldberg, Y. Dependency-based word embeddings. In *Proc. of 52nd Annual Mtg of Assoc. for Computational Linguistics* vol. 2, pp. 302-308. (2014).
8. Li, C., Wang, H., Zhang, Z., Sun, A., Ma, Z. . Topic modeling for short texts with auxiliary word embeddings. In Proc. of 39th Intl ACM SIGIR Conf on Research & Development in Information Retrieval, pp. 165-174. ACM. (2016).
9. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems, pp. 3111-3119. (2013)..
10. Pennington, J., Socher, R., Manning, C. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. (2014).
11. Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., Lazzara, E. H. (2015), Understanding and improving teamwork in organizations: A scientifically based practical guide. Human Resource Management, **54**, pp. 599-622. (2015).
12. Shalaby, W., Zadrozny, W., Jin, H. Beyond word embeddings: Learning entity and concept representations from large scale knowledge bases. Information Retrieval Journal, pp. 1-18. (2018).
13. Sinatra, A.M. Team models in the generalized intelligent framework for tutoring: 2018 update. In Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), pp. 169-173. (2018).
14. Sottilare, R., Holden, H., Brawner, K., Goldberg, B. Challenges and emerging concepts in the development of adaptive, computer-based tutoring systems for team training. In Proceedings of the Interservice/Industry Training Systems & Education Conference, Orlando, December 2011. (2011).
15. Sottilare, R., Holden, H., Goldberg, B., Brawner, K. The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottilare, R. (Eds.) Fundamental Issues in Defence Sim & Training. Ashgate. (2013).
16. Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., Gilbert, S. B. Towards a design process for adaptive instruction of Teams: A meta-analysis. International Journal of Artificial Intelligence in Education, **28**, pp. 225-264. (2018).
17. Sottilare, R., Graesser, A., Hu, X., Sinatra, A. (Eds.).*Design Recommendations for Intelligent Tutoring Systems: Volume 6 - Team Tutoring*. Orlando, FL: U.S. Army Research Laboratory. (2018).