TAC 2018 Systematic Review Information Extraction (SRIE) Track Guidelines

BACKGROUND

The National Toxicology Program (NTP), an interagency program headquartered at the National Institute of Environmental Health Sciences (NIEHS, part of the National Institutes of Health), and the Environmental Protection Agency (EPA) conduct systematic reviews of environmental agents to identify potential human health hazards. These reviews collect toxicity or health effects information on different chemicals from the published scientific literature including study details such as experimental protocols, animal models, and results. Because this information can vary widely from study to study, the systematic review serves a critical purpose by providing a transparent, standardized, multistep approach to identify, select, assess, and synthesize information for developing objective, evidence-based conclusions about potential chemical hazards. Furthermore, because research studies done at different points in time may reflect different standards in experimental protocols or reporting procedures, the systematic review approach serves to promote transparency and facilitate reproducibility of literature-based evaluations on environmental agents.

Some elements of information extraction in these systematic reviews are straightforward, such as identifying the species or sex of the experimental models. Others are more complex as publications may report multiple experiments with various exposures and doses and evaluate multiple endpoints. Authors may report experimental details using different units, different chemical names, and other variations in terminology. In addition, this information may be located in the text of the publication, or in a table, figure caption, or the figure itself. Currently, the information extracted in a systematic review is collected through a labor-intensive, manual process that is slow and often costly. NTP and EPA are interested in adopting automated processes for information extraction in systematic reviews of environmental chemicals. The application of this task is to develop automated tools that could improve the efficiency of systematic review information extraction to reduce completion time and labor-costs while maintaining quality and reproducibility. The results of this task will inform future NTP and EPA efforts aimed at systematic review automation, including subsequent challenges.

OBJECTIVE

The purpose of the Systematic Review Information Extraction (SRIE) track is to develop and evaluate Information Extraction (IE) approaches that can assist in the systematic reviews of environmental agents. This track will focus on IE of experimental design factors found in the Methods and Materials section ("methods section") of published studies of experimental animals exposed to environmental chemicals. The first goal of the track is to identify and annotate the experimental design factors. The second goal of the track is to identify relations between different experimental design factors and assign the factors into logical groups.

TASKS

The SRIE track has two tasks:

Task 1: Experimental design factors for the categories of exposure, animal group, dose group, and endpoint should be identified and the appropriate annotation tag applied.

Task 2: Relations between experimental design factors from Task 1 should be identified, the factors assigned to groups, and the appropriate annotation tag applied.

Task 2 builds on Task 1. Participants may choose to participate in only Task 1, or in both Task 1 and Task 2.

Task 1: Annotation of Mentions

The SRIE Annotation Guidelines released with the training data for Task 1 provide instructions regarding the types of factors within the Methods sections that should be annotated for the categories of exposure, animal group, dose group, and endpoint. All mentions of the factors should be annotated with the annotation tag.

Task 2: Annotation of Relations

The aim is to identify factors that are related within categories and across categories. Some relations will involve two factors and others multiple factors. The Annotation Guidelines released with the training data for Task 2 will identify the specific types of relations to annotate.

DATA

Training data will be released in two parts: first with the experimental design factors annotated and second with the relations annotated in a subset.

1. As training data for Task 1, mentions of experimental design factors within 100 Methods sections will be annotated using the BRAT annotation tool (http://brat.nlplab.org). Participants will be provided the 100 methods sections as text files (*.txt), BRAT-formatted annotation files (*.ann), and XML formatted annotation files (*.xml).
2. As training data for task 2, relations among experimental design factors will be annotated for a subset of the 100 methods sections. Participants will be provided the subset as text files, BRAT-formatted annotation files, and XML formatted annotation files.

For the test data set, participants will be provided ~300 Methods sections as text files to test their IE approach for identifying experimental design factors and their relations. Participants must submit an XML-formatted annotation file (*.xml) for each of the text files and will be evaluated on an unidentified subset hidden within the test data set.

Task 1 BRAT annotation file formats

Annotations for each document should be stored in a file with the same filename as the document except that the extension is ".ann" instead of ".txt".
Extracted mentions should be represented by the following columns:

Column 1: the letter T followed by a unique integer id
Column 2: the mention type, starting offset, ending offset +1

Column 3: the text associated with the mention

There is one row per mention and columns are tab delimited. The BRAT format is specified at http://brat.nlplab.org/standoff.html. See below for an example:

```
T1    GroupName 1081 1092     Female pups
T3    DoseRoute 1166 1178     sc injection
T4    Vehicle 1184 1192 corn oil
T5    TestArticle 1196 1199   GEN
T6    Dose 1201 1205    12.5
T9    DoseUnits 1217 1226     mg/kg/day
T10   TestArticlePurity 1228 1231   98%
T43   Endpoint 2103 2113;2118 2154    hippocampi examined six or seven
disector pairs
T60   TimeEndpointAssessed 2309 2338      4 hr after the last treatment
T61   Endpoint 2351 2363      body weights
```

Column 2 may contain multiple text spans indicated by multiple start and ending offsets; see T43 example above.

Task 1 XML annotation file formats

The Task 1 training annotations will be packaged in both the BRAT format and an XML format similar to that used by the TAC 2017 Adverse Drug Reaction (ADR) and TAC 2018 Drug-Drug Interaction (DDI) tracks:

```
<Doc name="PMCxxxxxxx">
<Text>
…
</Text>
<Mention id="T1" label="GroupName" span="1081 1092" str="Female pups"/>
<Mention id="T3" label="DoseRoute" span="1166 1178" str="sc injection"/>
<Mention id="T43" label="Endpoint" span="2103 2113; 2118 2154"
str="hippocampi examined six or seven disector pairs"/>

…
</Doc>
```

The document text is contained between the <Text>…</Text> tags, and spans refer to offsets from the first character following <Text>. The data in the XML files is converted straight from the BRAT output. Participants should submit XML formatted annotation files for Task 1.

Task 2 BRAT-formatted annotation file format

Annotations for each document should be stored in a file with the same filename as the document except that the extension is ".ann" instead of ".txt". Extracted mentions are represented as in Task 1, followed by group annotations. Group annotations should be represented by the following columns:

Column 1: the letter A followed by a unique integer id
Column 2: the text 'Group__' followed by the type of Group and an unique integer id for that Group type.
Column 3: id of a mention that is part of that group

There is one row per group- mention assignment and columns are tab delimited. See below for an example:

```
A1	GROUP__Animal-0 T68
A2	GROUP__Animal-0 T65
A3	GROUP__Dose-0 T6
A4	GROUP__Dose-1 T12
A5	GROUP__Endpoint-0 T108
```

Task 2 XML annotation file formats

The Task 2 training annotations will be packaged in both the BRAT format and an XML format (TBA). Participants should submit XML formatted annotation files for Task 2.

EVALUATION

For the test set, participants will be provided ~300 methods sections as text documents to test their IE approach for identifying experimental design factors and the assignment of factors into logical groups. Participants will be evaluated on an unidentified subset hidden within the test set. Participants will be expected to submit responses in the same xml format provided in the training data set.

The two tasks are:

Task 1	Precision/Recall/F1-measure on mention-level annotations, using both partial and exact matching. This task corresponds to Named Entity Recognition tasks. Primary Metric: micro-averaged F1 on exact and partial matching, overall and by mention type.
Task 2	Precision/Recall/F1-measure on assignment of entities to groups of entities. This task corresponds to Event Recognition tasks. Primary Metric: micro-averaged F1 on exact and partial matching, overall and by group type.

Evaluations metrics are computed by determining pairings between reference annotations and best matching predicted annotations (for both mention and group annotations). Best matching is based upon the similarity between each reference and predicted annotation. For entity level annotations, similarity is based on the overlap in spans, accounting for fragmented spans. For group annotations, similarity is based on overlap in number of (full or partial) matching mentions. Metrics will be reported for different similarity thresholds – for final comparison of methods, a similarity threshold of 40% will be used.

SUBMISSION

Participants will submit IE approach results on the entire set of unannotated Methods sections in the test set. Only the Methods sections for the unidentified subset hidden within the test set will be evaluated. Participants may submit results at any time between the release of the test set and the

deadline for submission of system results. Participants must freeze their system before downloading the test set, running their system, and submitting their results (i.e., participants are prohibited from tuning their system to the test set).

Participants are allowed 3 separate submissions per task. Submission that do not conform to the provided submission file format will be rejected without consideration. Instructions for submitting results will be announced on the SRIE track mailing list.

TIMELINE

June 1, 2018	Release of Task 1 training data
July 15, 2018	Deadline for registration for track participation
July 15, 2018	Release of Task 2 training data
August 15, 2018	Release of test set for Tasks 1 and 2
September 15, 2018	Deadline for submission of system results
October 1, 2018	Release of individual scores to participants
October 15, 2018	Deadline for short system descriptions and workshop presentation proposals
October 20, 2018	Notification of acceptance of presentation proposals
November 1, 2018	Deadline for system reports (workshop notebook version)
November 13-14, 2018	TAC 2018 Workshop in Gaithersburg, MD
February 15, 2019	Deadline for system reports (final proceedings version)

WEBSITE and MAILING LIST

SRIE 2018 guidelines, data, and tools will be distributed through the SRIE 2018 website at https://tac.nist.gov/2018/SRIE/. Announcements about SRIE will be sent to the track mailing list (tac-srie@lists.nist.gov), and subscribed individuals may also post questions to the mailing list. Participants should subscribe to the mailing list at https://groups.google.com/a/list.nist.gov/forum/#!forum/tac-srie.

ORGANIZING COMMITTEE

Charles Schmitt (charles.schmitt@nih.gov), Co-Track coordinator
Mary Wolfe (wolfe@niehs.nih.gov) Co-Track coordinator
Michelle Angrish (Angrish.Michelle@epa.gov)
Dina Demner-Fushman (consultant, ddemner@mail.nih.gov)
Kristan Markey (markey.kristan@epa.gov)
Andrew Rooney (rooney@niehs.nih.gov)
Michele Taylor (taylor.michelem@epa.gov)
Vickie Walker (vickie.walker@nih.gov)
Byron Wallace (consultant, byron@ccs.neu.edu)