

# ConceptGrid: Natural Language Processing for Intelligent Tutors

Shrenik Devasani, Stephen Gilbert, Gregory Aist, Stephen Blessing

## PROBLEM

Interpreting textual responses from a student by an [Intelligent Tutoring System \(ITS\)](#) is essential if it has to match the performance of a human tutor, since the use of language makes the learning process more natural.

Current [Natural Language Processing \(NLP\)](#) techniques do not give sufficient power to non-programmers to build ITSs that support NLP. In most cases, the NLP is left to expert developers or machine learning algorithms, and the tutor-author focuses on just tutoring strategies.

## SOLUTION

ConceptGrid is a domain-independent [authoring tool](#) that helps non-programmers develop ITSs that perform NLP.

The crucial part about developing an authoring tool for non-programmers is managing the trade-off between its ease of use and expressiveness. ConceptGrid's ease of use and expressiveness lies between that of simple word matching approaches and complex machine learning algorithms.

The author develops the NLP component of a tutor by breaking down an expected response into specific [concepts](#). Templates are designed that describe a set of required concepts and incorrect concepts. There is a one-to-many relationship between concepts and templates.

A [template](#) consists of one or more atomic [checktypes](#) that are based on algorithms and distance measures, such as Levenshtein, Soundex and Stemmer. Two more checktypes, "Any" (matches a range of words), and "Not" (checks for negation) help make a template more expressive.

## INTERFACE

ConceptGrid uses a web-based [lattice-style table driven interface](#) that allows the author to design templates.

For example, if the correct response to a question is "Reject the null hypothesis", the template shown in the image below helps model the concept "Rejection-Correct" that refers to the student correctly mentioning that the null hypothesis must be rejected.

+	X	+	X	+	X	+	X	+	X	+
Not	2	<-	Levenshtein	1	Any	0	2	Levenshtein	1	Soundex
X	fail		reject					null		hypothesis
X	not									
+										

Once all the concepts have been designed, the author can create a ternary truth table called the [Feedback Table](#), that helps provide customized just-in-time feedback to a student based on the truth values of the concepts: "true," "false," or "don't care."

Rejection Correct	Rejection Incorrect	Significance Correct	Significance Incorrect	Feedback
—	—	—	✓	You haven't correctly mentioned the significance of the results.
✗	✗	—	—	Should the null hypothesis be accepted or rejected?
✓	—	—	✓	Your statements of rejection and significance are inconsistent.

## EVALUATION

- As a part of the xSTAT project, [41 students](#) solved a total of [233 instances](#) of six college-level statistics problems.
- Each problem had a question that asked for the conclusion of a statistics test. The corpus thus created consisted of [554 unique responses](#).

✓ What is the obtained p value?  
0.05

✗ What do you conclude?  
There is a significant difference in sleep efficiency scores between students who are depressed and the population mean.

ConceptGrid

You have correctly calculated the p-value for the test. But, are you sure that there is a significant difference between the 2 conditions?

## RESULTS

- A course instructor manually scored the corpus based on the presence or absence of [8 concepts](#).
- A tutor-author developed [22 templates](#) across all six problems, using ConceptGrid.
- ConceptGrid had a [concurrency](#) of over 87% with the human grader.

### Results of classification by ConceptGrid:

Concept	False +	False -	Accuracy	Kappa
Rejection-Correct	1	34	0.9368	0.8657
Rejection-Incorrect	6	5	0.9801	0.9217
Significance-Correct	1	7	0.9856	0.9662
Significance-Incorrect	12	1	0.9765	0.8890
Ind-Variable-Mention	1	3	0.9928	0.9853
Dep-Variable-Mention	4	3	0.9874	0.9733
Conclusion-Correct	0	24	0.9567	0.8614
Conclusion-Incorrect	6	0	0.9892	0.9727

## DISCUSSION

ConceptGrid's tries to "understand" responses by looking for [smaller concepts](#) and [key phrases](#), rather than literal string matching. The interface provides an intuitive way to design templates and provide feedback.

It adopts a [non-structural](#) approach focused on words and numerical analysis. This makes it domain-independent and simple for non-programmers to use.

ConceptGrid can be very effective in domains such as statistics, where the student responses are expected to follow a [general pattern](#).

This work is supported with funding from the AFOSR.