

EqFix: Fixing \LaTeX Equation Errors by Examples^{*}

Fengmin Zhu^{1,4}[0000–0003–4219–0837] ^{**} and Fei He^(✉)1,2,3[0000–0002–4266–875X]

¹ School of Software, Tsinghua University

² Key Laboratory for Information System Security, MoE, China

³ Beijing National Research Center for Information Science and Technology

⁴ Max Planck Institute for Software Systems

hefei@tsinghua.edu.cn

Abstract. \LaTeX is a widely-used document preparation system. Its powerful ability in mathematical equation editing is perhaps the main reason for its popularity in academia. Sometimes, however, even an expert user may spend much time fixing an erroneous equation. In this paper, we present EqFix, a synthesis-based repairing system for \LaTeX equations. It employs a set of fixing rules and can suggest possible repairs for common errors in \LaTeX equations. A domain-specific language is proposed for formally expressing the fixing rules. The fixing rules can be automatically synthesized from a set of input-output examples. An extension of relaxers is also introduced to enhance the practicality of EqFix. We evaluate EqFix on real-world examples and find that it can synthesize rules with high generalization ability. Compared with a state-of-the-art string transformation synthesizer, EqFix solved 37% more cases and spent less than half of their synthesis time.

Keywords: Domain-specific languages · Program synthesis · Program repair · Programming by examples.

1 Introduction

\LaTeX is a text-based document preparation system widely used in academia to publish and communicate scientific documents. The powerful typesetting of mathematical equations makes it a universal syntax for expressing mathematical equations. This syntax has been integrated into text-based markup languages like Markdown⁵, and WYSIWYG (i.e., “What you see is what you get”) document processors like MS Word⁶. Even on the web, one can display \LaTeX equations beautifully by MathJax⁷.

^{*} This work was supported in part by the National Natural Science Foundation of China (No. 62072267 and No. 62021002) and the National Key Research and Development Program of China (No. 2018YFB1308601).

^{**} Early revisions of this work were done when this author was in Tsinghua University.

⁵ <https://daringfireball.net/projects/markdown/>

⁶ <https://products.office.com/en-us/word>

⁷ <https://www.mathjax.org>

Since the syntax of \LaTeX equations is quite *complex, non-expert* users may find it challenging to use. For example, one may expect “ x^{10} ” by typing “ $\$x^{10}\$$ ”. This equation indeed compiles; however, its actual output is “ x^{10} ”, which goes against the user’s expectation. When such an error occurs, one may resort to online help forums. However, this process is *never trivial*. First, the user has to provide several keywords (for searching) that well describe the error. Second, even if the user is fortunate to obtain some solutions, they may not necessarily work for the user’s problem—the user still has to adapt the answers to that specific problem. The whole process—especially the adaptation of the solutions to one’s own problem—requires not only intelligence but also patience.

Programming by examples (PBE) is believed to be a revolutionary technique for end-user programming [8,7,10,12,1]. In recent decades, PBE has been adopted in the area of *program repair* [17,14]. NoFAQ [2] is a tool that employs *error messages* to assist the PBE-based repairing. This tool aims to fix common errors in Unix commands, from an input-output example that consists of an erroneous Unix command, an error message prompted by Shell, and a rectified command specified by experts. Note that the error messages prompted by Unix Shell are usually instantiated from a set of predefined templates; the stored information in the messages can thus be easily extracted by patterns.

Inspired by NoFAQ and other PBE techniques, we present EqFix, a system for automatically fixing erroneous \LaTeX equations by examples. Note that NoFAQ cannot be applied to our problem setting for two reasons: First, an equation error is not necessarily a compilation error, such as “ $\$x^{10}\$$ ” indeed compiles but produces an unexpected result. For such errors, one needs to specify the error message on their own. Thus, we cannot assume this message has a fixed structure (i.e., is instantiated from a template) as in NoFAQ. Second, the Unix command can be directly *tokenized* into a sequence of strings (using whitespaces as the delimiters), which makes it straightforward (by comparing the tokenized strings in turn) to locate the error position in the text of the Unix command. However, it is not the same case for equation text. Instead, we have to collaborate the corresponding *error message* to tokenize an equation.

To the best of our knowledge, EqFix is the first attempt at equation repair using PBE techniques. Novice users can use it to automatically fix common errors in \LaTeX equations; expert users can contribute corrections for erroneous equations. We design a *domain-specific language* (DSL) for formally defining the fixing rules for erroneous \LaTeX equations (§ 3). Intuitively, a fixing rule consists of an *error pattern*, which specifies what error messages this rule is applicable, and a *transformer* which performs the actual fixing via string transformation. We propose an algorithm for synthesizing fixing rules expressed by our DSL from input-output examples (§ 4). We also introduce *relaxers* to describe the *generalization* of equation patterns, with which the search space relating to the faulty parts is expanded so that we can handle more repairing problems.

We evaluated EqFix on a dataset containing 89 groups of real-world examples (§ 5). Note that NoFAQ is limited to repairing buggy Unix commands; we instead took the state-of-the-art PBE tool FlashFill as our baseline. We selected the

longest example in each group as our test case. We found that EqFix solved 72 (80.9%) of the test cases in less time, whereas FlashFill solved only 39 (43.82%). Our prototype implementation and experiment artifact are publicly available: <https://github.com/thufv/EqFix>.

The main technical contributions of this paper are summarized as follows:

- We present EqFix, a PBE-based system for automatically fixing erroneous L^AT_EX equations.
- We design a DSL for formally expressing fixing rules. We rely on equation patterns to extract and transfer relevant information between error messages and equations. The patterns can be generalized by relaxers when necessary.
- We conducted experiments on real-world examples. Results reveal the high effectiveness and applicability of our approach.

2 EqFix by Examples

We use several real-world examples (in Table 1) to showcase how EqFix repairs L^AT_EX equations. All of the examples were extracted from an online L^AT_EX forum⁸. To be clear and short, we neglect the unchanged substrings of long equations. Each example consists of three components – the input equation eq , the error message err , and the fixed equation fix . For convenience of reference, we number these examples from 1 to 8 and refer their components as eq_i , err_i , fix_i , for $1 \leq i \leq 8$, respectively. The L^AT_EX output on each equation (if it compiles) is displayed below the equation text.

Examples #1 – #4 present a scene where a user expects a superscript but forgets to parenthesize the superscript expression. As shown in #1, given the input equation “ x^{10} ” (eq_1), L^AT_EX treats only “1” but not the entire number “10” as the superscript. In this way, it outputs “ x^10 ”, which is against the user’s intent. Note that the input equation eq_1 itself is syntactically correct because the L^AT_EX compiler did not report any error. In this way, the user must specify an error message by hand to express their intent. To express the error type conveniently, a set of predefined *keywords* (Table 2 presents a selected subset) are provided for selection. In a future direction, via natural language processing, we may simply accept a natural language sentence as the error message for even better practicality. Then, the user needs to point out a substring of the erroneous input equation to show the error location. The error message is a combination of the keywords and substring of the input equation. Here, in err_1 , “**superscript**” is a predefined keyword indicating some substring of the input equation is expected to be the superscript, and “10” gives the error location. Then, an expert may fix⁹ the input equation as “ $x^{\{10\}}$ ”, i.e., surrounding “10” with a pair of curly brackets. The three components, i.e., (eq_1, err_1, fix_1) , compose an *input-output example*, with which we can synthesize a (*fixing*) rule. Each rule consists of an *error pattern* for matching the error message and a *transformer* that will

⁸ <https://tex.stackexchange.com>

⁹ As another option, we may get the fix by online search.

Table 1: Motivating examples.

# eq	err	fix
1 x^{10}	superscript 10	x^{10}
2 $y^{123} + x$	superscript 123	$y^{123} + x$
3 $f^{(k)}$	superscript (k)	$f^{(k)}$
4 $y = x + \dots + x^{10}$	superscript 10	$y = x + \dots + x^{10}$
5 $\{1, 2, 3\}$	Missing } inserted	$\{1, 2, 3\}$
6 $S = \{x_1, \dots, x_n\}$	Missing } inserted	$S = x_1, \dots, x_n$
7 2^x	Command ^ invalid in math mode	2^x
8 $\sum_{i=1}^N t_i$	Command ^ invalid in math mode	$\sum_{i=1}^N t_i$

be applied to the input equation to produce a fix. Intuitively, the underlying fixing strategy of this rule would be “surrounding the superscript with a pair of curly brackets”.

EqFix can switch between the *training mode* for synthesizing rules and the *applying mode* for repairing erroneous equations, based on a *rule library* that saves all the learned rules so far. In the training mode, EqFix takes user-given examples (typically by expert users) as input. It first searches in its rule library to obtain a rule that can be *refined* to be *consistent* with the new examples. For instance, the example #2 can be added by refining the fixing rule synthesized merely by #1. If it is not the case (for instance, consider the examples #5 to #8), a new rule is synthesized and the rule library gets enlarged.

In the applying mode, EqFix attempts to solve an *equation repair problem*—an erroneous equation together with an error message—typically provided by an end user. To do so, it searches in its rule library for all applicable rules, i.e., those whose error patterns can match against the error message, and attempts to apply them (the transformer of the rule) to the input equation. For instance, the rule synthesized from examples #1 and #2 is applicable to equation repair problems #3 and #4: applying this rule on (eq_3, err_3) and (eq_4, err_4) gives fix_3 and fix_4 respectively. Since there can be more than one applicable rule, users are asked to review the suggested fixes and approve one that meets the intent. If no rule is applicable, or all suggested fixes are rejected by the user, EqFix fails on this equation repair problem. In that situation, we expect an expert user to figure out a correction, which, in association with the erroneous equation and the

Table 2: Selected keywords supported by EqFix.

Keywords	Interpretations
superscript	expected as a superscript
subscript	expected as a subscript
set	expected as a set
function	expected as a math function/operator
greek letter	expected as a greek letter
fraction numerator	expected as the numerator of a fraction
fraction denominator	expected as the denominator of a fraction
operator sum	expected as a sum operator
operator product	expected as a product operator
long arrow	expected as a long arrow

error message, forms a new example for synthesis under the training mode. The newly synthesized rule will be recorded in the rule library so that it can apply (under the applying mode) to future equation repair problems in this category.

Sometimes, an erroneous equation contains multiple errors. One needs to interact with EqFix in multiple rounds to fix them all. The user feeds the erroneous equation together with one of the error messages in the initial round and iteratively corrects the other errors using the fixed equation of the last round.

The rest examples in Table 1 showcase two L^AT_EX_{compile} errors: unmatched brackets (#5 – #6) and invalid superscript operator (#7 – #8). The error messages prompted by the L^AT_EX_{compile} are instantiated from some templates defined by L^AT_EX. Both compiler-prompted and keywords-based error messages are handled in a unified way (we will explain that in § 3.1). Back to the examples, a possible correction for #5 (#6 is similar) suggested by an expert is “ $\${1,2,3}\$$ ”, which inserts the missing right curly bracket (‘}’) at the end of the equation. A possible correction for #7 (and #8) is to use ‘^’ in place of the erroneous “\^”. Note that #8 is more complicated than #7, while it can be automatically fixed using the rule synthesized from #7.

In summary, EqFix facilitates an automated approach for fixing common errors in editing L^AT_EX equations. Our approach is rule-based (§ 3) and the synthesis by input-output examples (§ 4) is automated. One benefit of our system is that we may collect many examples and train a set of fixing rules from them in advance that covers many common problems end users meet. Another benefit is that the manual efforts of adapting the searched correction to their cases, which might be the most challenging part to end users, are saved.

3 Rules in EqFix

EqFix is a rule-based system. Rules are formally defined by a DSL as shown in Fig. 1. Each *rule* \mathcal{R} is a pair $\langle EP, \mathcal{T} \rangle$, where *EP* is an *error pattern* describing the template of the error message and which problem-specific information needs to

(Fixing) rule $\mathcal{R} ::= \langle EP, \mathcal{T} \rangle$
 Transformer $\mathcal{T} ::= \{v_1 \mapsto \tau_1, \dots, v_k \mapsto \tau_k\}$
 Error pattern $EP ::= [M_1, \dots, M_k]$
 Matcher $M ::= s \mid v$
 Equation pattern $P ::= [M_1, \dots, M_k]$

Fig. 1: Syntax of fixing rules.

be extracted from that message, and \mathcal{T} is a *transformer* specifying the required transformation on the erroneous equation to fix the equation repair problem.

3.1 Error Pattern

An error message either comes from the L^AT_EX compiler (e.g., #5 – #8 of Table 1) or the user (e.g., #1 – #4). For the latter case, we assume the error message starts with one or more predefined *keywords* (as in Table 2) that mention the error type and then followed by a substring of the erroneous equation which locates the error. In either case, we represent the error message as a natural language sentence that can be split into a list of *tokens* $[e_1, \dots, e_k]$ by delimiters (whitespaces, commas, etc.). EqFix is unaware of the resource of the error messages and employs a unified *error pattern* to match against them. Users are allowed to customize their keywords because EqFix regards them as normal tokens.

An *error pattern* $EP = [EM_1, \dots, EM_k]$ contains a list of *matchers*, where each of them is either: (1) a string matcher s that only matches against s itself, or (2) a variable matcher v that matches against any token and binds the matched token to v . An error message $[e_1, \dots, e_{k'}]$ matches EP if they have the same length ($k = k'$), and that every token e_i ($1 \leq i \leq k$) matches the corresponding matcher M_i . The matching result (if succeeds) is a mapping from variables to the bound string values $\{v_1 \mapsto s_1, \dots, v_k \mapsto s_k\}$.

Example 1. Consider $err_2 = [\text{“superscript”}, \text{“123”}]$ from Table 1. Let $EP_1 \triangleq [\text{“superscript”}, v_1]$ be an error pattern. Matching err_2 against EP_1 succeeds with $\{v_1 \mapsto \text{“123”}\}$, i.e., “123” is bound to v_1 .

3.2 Equation Pattern

Unlike an error message, an equation text usually involves complicated syntax and thus cannot be directly tokenized by commonly seen delimiters. To extract the problem-specific information from an equation (e.g., to find the cause of the error), we propose the notion of an *equation pattern*.

An *equation pattern* $P = [M_1, \dots, M_k]$ consists of a list of matchers. Especially, the string and variable matchers in P must appear *alternately*, that is, if M_i is a string, then M_{i+1} must be a variable and vice versa. Intuitively, the

string matchers in an equation pattern are indeed used as “delimiters” to tokenize the equation into a list of “tokens”. Each “token” of the equation matches a variable matcher and may convey some *problem-specific* information which may be useful later in generating a corrected equation. Oppositely, if we allow consecutive variable matchers to appear in an equation pattern, the split would be *ambiguous*. For instance, consider a string “`alpha`” and an equation pattern $[v_1, v_2]$, we could either let v_1 match against “`a`” (and v_2 match against “`lpha`”), or let v_1 match against “`al`” (and v_2 match against “`pha`”), etc.

Pattern Matching Given that string and variable matchers appear alternately, we pattern match an equation pattern P against an equation eq by simply locating the occurrences of the string matchers in eq (failure if we cannot)—the variable matchers then match against the substrings in between. For example, matching $[v_1, \text{“foo”}, v_2]$ against “`(foo)`” yields the bindings $\{v_1 \mapsto \text{“(”}, v_2 \mapsto \text{“)”}\}$, because the equation is split into three parts “`(` · “`foo`” · “`)`”.

Pattern Instantiation Equation patterns can be regarded as “templates” of equation text where the variable matchers are “placeholders”. Thus it is natural to define pattern *instantiation*—the reverse of pattern matching—to obtain a (concrete) equation by replacing the variable matchers with the bound strings.

Example 2. Consider example #2 of Table 1, where $eq_2 = \text{“}\$y^123+x\$”$. Let $P_2 \triangleq [\text{“}\$y^”, v_1, \text{“}+x\$”}]$ be an equation pattern. Matching eq_2 against P_2 gives $\sigma = \{v_1 \mapsto \text{“123”}\}$. Further, instantiating P_2 with σ gives back eq_2 .

Pattern Generation In EqFix, equation patterns are not explicitly presented in the rule. They are only *intermediate* during rule application. To construct an equation pattern P from an erroneous equation eq with the matching result $\sigma = \{v_1 \mapsto s_1, \dots, v_k \mapsto s_k\}$ from an error pattern, we substitute all the occurrences of s_1, \dots, s_k in eq with v_1, \dots, v_k , respectively.

Example 3. Given $\sigma = \{v_1 \mapsto \text{“123”}\}$ and $eq_2 = \text{“}\$y^123+x\$”$, applying the above process yields $[\text{“}\$y^”, v_1, \text{“}+x\$”}]$.

3.3 Transformer

Our DSL achieves an underlying repairing strategy via string transformation. Since string matchers express the *problem-unspecific* information, the substrings in the erroneous equation matched by them should be kept in the corrected equation. The substrings matched by variable matchers, on the other hand, need to be transformed by *string transformers*—functions that map a string into a new one. The string transformer we employ in EqFix is expressed by a variant of FlashFill’s DSL [7] but possesses a more restricted syntax for better efficiency. Technical details can be found in our extended version [29]. In EqFix, we define a *transformer*—a mapping from the variable matchers into string transformers—to collect all necessary string transformations.

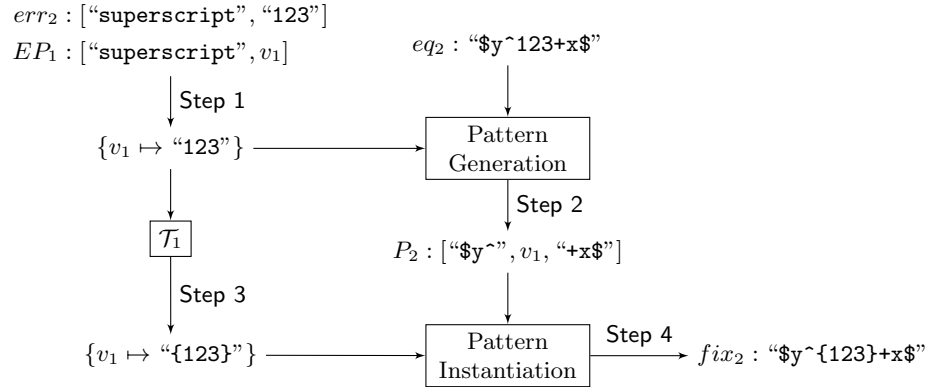


Fig. 2: Application of the rule $\mathcal{R}_1 = \langle EP_1, \mathcal{T}_1 \rangle$ to the input (eq_2, err_2) .

Example 4. Let τ_1 be a string transformer that inserts a pair of curly parentheses surrounding the input. Let $\mathcal{T}_1 \triangleq \{v_1 \mapsto \tau_1\}$ be a transformer. Applying \mathcal{T}_1 to $\sigma = \{v_1 \mapsto \text{"123"}\}$ yields $\sigma' = \{v_1 \mapsto \text{"{123}"}\}$.

3.4 Rule Application

Putting the above operations together, we now present how a rule $\mathcal{R} = \langle EP, \mathcal{T} \rangle$ is applied to an equation repair problem (eq, err) :

1. match EP against err to extract problem-specific information recorded in a mapping σ ;
2. generate an equation pattern P (from σ), regarded as an “template” of eq ;
3. perform the underlying repairing strategy expressed by the set of string transformers in \mathcal{T} on σ to obtain a new σ' ;
4. obtain the corrected equation by instantiating P with σ' .

Example 5. Let rule $\mathcal{R}_1 \triangleq \langle EP_1, \mathcal{T}_1 \rangle$, where EP_1 is defined in Example 1 and \mathcal{T}_1 is defined in Example 4. Following the above steps, we apply \mathcal{R}_1 to the equation repair problem (eq_2, err_2) (depicted by Fig. 2):

1. matching EP_1 against err_2 gives σ (as in Example 1);
2. generate P_2 from eq_2 and σ (as in Example 3);
3. transform σ into σ' by \mathcal{T}_1 (as in Example 4);
4. instantiating P with σ' gives fix_2 (as in Example 2).

4 Rule Synthesis

The rule synthesis algorithm takes a set of input-output examples \mathcal{E} as the specification, and generates a fixing rule $\mathcal{R} = \langle EP, \mathcal{T} \rangle$ consistent with the examples. The synthesis consists of two passes: (1) an error pattern EP is synthesized from the examples \mathcal{E} , and (2) a transformer \mathcal{T} is synthesized from \mathcal{E} and EP .

4.1 Synthesizing Error Patterns

Given an error message $err = [e_1, \dots, e_k]$, our problem is to generate an error pattern EP (of the same length) that matches against err . To achieve this goal, a naive pattern $EP_{\perp} = [e_1, \dots, e_k]$, the error message itself, is apparently a solution. However, it is so *restricted* that only this error message can match it. Oppositely, another naive pattern $EP_{\top} = [v_1, \dots, v_k]$ is too *general* and can be matched with any error message with k tokens.

To synthesize an error pattern that is neither too restricted nor too general, we start with EP_{\perp} , and for each string in EP_{\perp} , replace it with a fresh variable; if it also occurs in either the input equation eq or the output equation fix . Such a replacement makes the error pattern more general. Realizing that the error message is usually related to either the input equation by telling why it is erroneous or the output equation by explaining how to repair it, the introduced variables, in either case, will capture such important information.

4.2 Synthesizing Transformers

The essential problem of synthesizing a transformer $\{v_1 \mapsto \tau_1, \dots, v_k \mapsto \tau_k\}$ is to synthesize the underlying string transformers τ_1, \dots, τ_k . This problem has been well-studied in previous literature, and a well-known approach could be the PBE approach invented by FlashFill [7]. To adopt their approach in our setting, we must extract a set of *input-output string examples*, each is a pair (s, s') packed the input string s and the expected output string s' , as the specification for synthesizing each string transformer τ_i .

Let us first consider the situation where the generated pattern (using the pattern generalization process mentioned in § 3.2) matches against both the input and output equations for all examples \mathcal{E} . This condition implies that we can always compute two mappings by matching the generated pattern against the input and output equation. Thus, to synthesize τ_i , we are able to extract an input-output string example $(\sigma(v_i), \sigma'(v_i))$ for each example in \mathcal{E} (so the complete specification is their union), where σ and σ' are the two mappings computed as above.

Example 6. Given examples #1 and #2 from Table 1, to synthesize a transformer τ_1 , we extract (“10”, “{10}”) from example #1, and (“123”, “{123}”) from example #2 (see Example 3 for the generated pattern). Thus, the complete specification for synthesizing τ_1 is $\varphi_1 = \{(\text{“10”}, \text{“\{10\}”}), (\text{“123”}, \text{“\{123\}”})\}$.

Suppose the generated pattern only matches against the input equation but not the output; we must generalize this pattern so that it matches against the output equation. The generalization process is an extension to EqFix, and we will discuss it later in § 4.3. Once this is done, the synthesis method we have just introduced works again. So far, we have adequate mechanism to synthesize a rule from example #1 of Table 1 (depicted in Fig. 3):

1. synthesize an error pattern EP_1 by comparing err_1 with eq_1 and fix_1 , respectively, $\sigma_1 = \{v_1 \mapsto \text{“10”}\}$ records the values of the matched variable;

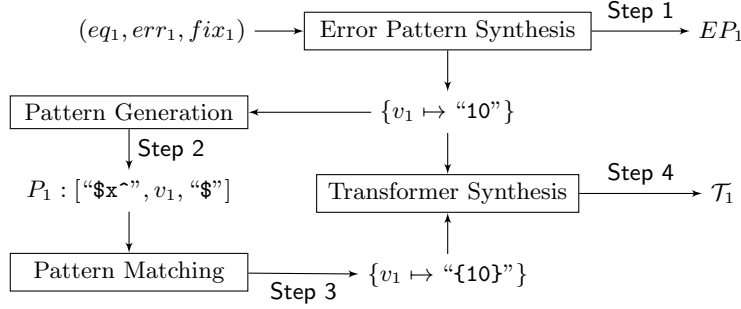


Fig. 3: Synthesis a rule $\mathcal{R}_1 = \langle EP_1, \mathcal{T}_1 \rangle$ by example (eq_1, err_1, fix_1) .

2. generate the equation pattern P_1 ;
3. match P_1 against fix_1 , yielding $\sigma'_1 = \{v_1 \mapsto “\{10\}”\}$;
4. synthesize τ_1 from $\{ (“10”, “\{10\}”)\}$, which gives rise to the transformer $\mathcal{T}_1 = \{v_1 \mapsto \tau_1\}$.

4.3 Extension: Pattern Generalization via Lazy Relaxation

Let us study some examples to get a sense of how to generalize patterns.

Example 7. Consider example #5 of Table 1, where

- $eq_5 = “\{1,2,3\}”$,
- $fix_5 = “\{1,2,3\}”$, and
- $err_5 = “Missing } inserted”$.

Let $EP_2 \triangleq [“Missing”, v_1, “inserted”]$ be the error pattern. Matching EP_2 against err_5 gives $\sigma_5 = \{v_1 \mapsto “}”\}$. From eq_5 and σ_5 , an equation pattern $P_5 = [“\{1,2,3\}”]$ is generated. We see that P_5 cannot be matched by fix_5 . However, a more general pattern such as $[v]$ (where v is a fresh variable) matches with fix_5 .

Example 8. Consider $eq = s_1 \cdot s_2 \cdot s_3 \cdot s_4$, $fix = s_1 \cdot s'_2 \cdot s'_3 \cdot s'_4$, and $\sigma = \{v_1 \mapsto s_2, v_2 \mapsto s_4\}$. Note that s_2, s_3 and s_4 are all modified in fix compared to eq , however s_3 is not bound to a variable in σ . From eq and σ , an equation pattern $[s_1, v_1, s_3, v_2]$ is generated, which apparently does not match against fix . However, a more general pattern like $[s_1, v]$ (v is a fresh variable) can match fix , where v is matched against $s'_2 \cdot s'_3 \cdot s'_4$.

We learn from the examples that if the generated equation pattern cannot match the output equation, we can always replace several string matchers with fresh variables until it matches against the output equation – we call this process *pattern relaxation*. As the relaxation goes on, the pattern becomes more and more general. In the worst case, it gives $P_\top = [v]$ that consists of a single variable v and can be matched by any string (as in Example 7). To find a relaxed pattern that is as strict as possible, the relaxation should be *lazy*.

Lazy Relaxation Process Let P be an equation pattern, and s be a string (i.e., the output equation). As a special case, if P is a constant (i.e., contains no variables), P_{\top} is the only possible relaxed pattern. Otherwise, since P cannot match against s , there must be some string s' in P that is not a substring of s . Three kinds of relaxations are performed on P depending on the relative position (left, right, or binary) of s' in P :

(left) If s' appears at the beginning of P , we replace the subpattern that consists of s' and the variable V followed by it, with a fresh variable $\text{LVar}(V)$.

(right) If s' appears in the end of P , we replace the subpattern that consists of s' and the variable V before it, with a fresh variable $\text{RVar}(V)$.

(binary) If s' appears in the middle of P , we replace the subpattern that consists of s' and the adjacent variables V_1 and V_2 , with a fresh variable $\text{BVar}(V_1, V_2)$.

The above repeats until the current pattern already matches against s .

Example 9. Pattern P_5 in Example 7 is relaxed to P_{\top} .

Example 10. Pattern $[s_1, v_1, s_3, v_2]$ in Example 8 is relaxed to $[s_1, \text{BVar}(v_1, v_2)]$.

Example 11. Consider a pattern $[s_1, v_2, s_3, v_4, s_5, v_6, s_7]$ and an output equation $s'_1 \cdot s_2 \cdot s_3 \cdot s_4 \cdot s'_5 \cdot s_6 \cdot s'_7$. It takes several steps to obtain a relaxed result. In the following, we highlight the relaxed subpattern with an underline and annotate the unmatched string with an asterisk:

$$\begin{aligned} [\underline{s_1^*}, v_2, s_3, v_4, s_5, v_6, s_7] &\rightarrow [\text{LVar}(v_2), s_3, \underline{v_4, s_5^*, v_6}, s_7] \\ &\rightarrow [\text{LVar}(v_2), s_3, \text{BVar}(v_4, v_6), s_7^*] \\ &\rightarrow [\text{LVar}(v_2), s_3, \text{RVar}(\text{BVar}(v_4, v_6))] \end{aligned}$$

Relaxers and Synthesis We extend the rule DSL (Fig. 1) to include *relaxers* that syntactically encode the three kinds of relaxations, with `id` for no relaxation:

$$\begin{aligned} \text{Rule } \mathcal{R} &::= \langle EP, \{r_1, \dots, r_k\}, \mathcal{T} \rangle \\ \text{Relaxer } r &::= \text{id}(v) \mid \text{LRelax}(r) \mid \text{RRelax}(r) \mid \text{BRelax}(r_1, r_2) \\ \text{Variable } V &::= v \mid \text{LVar}(V) \mid \text{RVar}(V) \mid \text{BVar}(V_1, V_2) \end{aligned}$$

In applying a rule with relaxers, one or more subpatterns of the generated equation pattern will be substituted with fresh variables according to the relaxers. For synthesis, the above lazy relaxation process is performed, and the relaxations that have been applied are recorded as corresponding relaxers. For more technical details, please refer to our extended version [29].

5 Evaluation

We prototyped the proposed approach as a tool EqFix, written in a combination of F# and C#, running on the .NET core platform. Rule application, synthesis

algorithms, and relaxer extensions were built following the approaches proposed in the paper. We developed the synthesizer for string transformation under the PROSE framework (proposed in [22]), in which we specified the syntax and semantics of our DSL, a set of witness functions for guiding PROSE’s synthesis engine, and a bunch of scoring functions for ranking candidate programs.

To measure the performance of EqFix, we conducted an experimental comparison with FlashFill [7], a state-of-the-art synthesizer for string manipulation, on a dataset that consists of 89 input-output *example groups* collected from the web (online help forums, tutorials, and technical blogs), each reveals one type of common mistake that users make, such as mismatch of delimiters and misuse of commands. Unlike the machine learning approaches, PBE techniques usually only need a few (2 – 5) examples. The lengths of the erroneous equations vary from 5 to 166, with an average of 18.

5.1 Experimental Setup

The baseline tool FlashFill was initially designed for string manipulation in spreadsheets (such as Excel), so an input-output example comprises a column of strings as input and a single string as output. To adapt FlashFill to our problem domain, we regarded an erroneous equation and an error message as two *indistinguished* input columns.

Another difference between FlashFill and EqFix is that FlashFill does not maintain a rule library. To make a fair comparison, we made the following adaption to avoid the usage of rule libraries: both tools were tested on an equation repair problem immediately after the rules were synthesized using the examples under the same example group. Since the number of consistent string transformers with a given specification is usually multiple, our synthesis algorithm will produce multiple candidate rules as well for one group of input-output examples. The candidate rules were ranked with heuristics, and we only attempted the top-10 candidate rules for each test case for fixing. If any rule gives a fixed equation that equals the expected correction of that test case, the test case is said “solved” (otherwise “failed”). In each example group, we left the one with the longest erroneous equation as the test case and the others as training examples. We set four training configurations C_1, \dots, C_4 , where C_i ($i = 1, \dots, 4$) means the first i shortest (by the length of the erroneous equation) examples in the training set are used for synthesis. There were, in total, 356 training runs of EqFix (and also for FlashFill).

The experiments were conducted on an Intel(R) Core(TM) i5 laptop with 2.3 GHz CPU and 8 GB memory, running Mac OS 11.6 and .NET core 2.2.207.

5.2 Results

The overall number of solved test cases is presented in Fig. 4. EqFix outperformed FlashFill under all configurations. When training with only one example (C_1), EqFix solved 67 (75.2%) test cases, whereas FlashFill failed to solve any. Both EqFix and FlashFill performed better when more training examples were

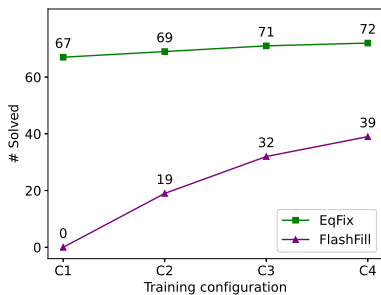


Fig. 4: Overall number of solved test cases.

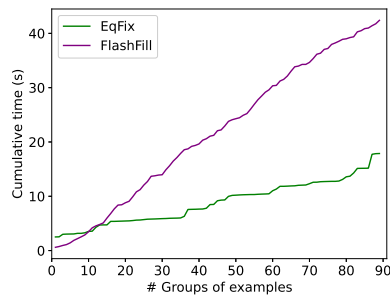


Fig. 5: Cumulative synthesis time (s) with an increasing number of training example groups (C4).

given—this is well-understood, as more training examples eliminate spurious rules and lead to more general rules. With the entire training set (C4), EqFix solved 72 (80.9%) while FlashFill solved less than half of the test cases. Therefore, EqFix has a more vital learning ability in our problem domain.

Impact of Ranking To understand how ranking affects the ability to solve the test cases, we list the numbers of attempted rules¹⁰ in Table 3. A cross mark “**X**” indicates that no (top-10) rules produced the expected fix. Among the solved test cases by EqFix, at most 5 rules were attempted (#81), and 57 – 68 test cases were solved by the top-ranked rule. In contrast, FlashFill attempted at most 8 rules (#48), and 0 – 33 test cases were solved by the top-ranked rule. We also recognize that providing more training examples helps to decrease the number of attempts, e.g., in #4 and #82.

Generalization Ability Thanks to the dynamic generation of equation patterns, the rules learned by EqFix are *insensitive* to where errors locate. In example #1 of Table 1, the erroneous part “10” appears at the end of the input equation, whereas in example #2, “123” appears before “+x” in the input equation. Although their positions are distinct, the rule synthesized by example #1 is general enough to fix problem #2, as the generated equation pattern P_2 (in Fig. 2) replaces “123” with v_1 . In contrast, the rules learned by FlashFill are less general—in many cases, they are *position-sensitive* because FlashFill does not support extracting problem-specific information from error messages, which is a major difference between the two. As a result, the rule learned from #1 by FlashFill cannot generalize to solve #2 while EqFix can.

Efficiency Fig. 5 depicts the cumulative synthesis time when the number of training example groups increases under C4. The average time was 201 ms for EqFix and 476 ms for FlashFill. EqFix spent less time on 78 out of 89 runs.

¹⁰ Rules were attempted in the order of the rank list.

Relaxer Extension We also noticed that in 17 (19.1%) of the 89 synthesized rules under configuration C4, relaxers exhibited in the synthesized rule. This reveals that the relaxer extension is necessary and improves our tool’s practicality.

Failure Cases EqFix failed on 17 test cases under C4. Manually inspecting these cases, we classified the cause of failure into three categories:

- Inconsistent examples (7 cases): The provided examples are *inconsistent* with each other, so our synthesizer failed to give any consistent rule.
- Insufficient error message (6 cases): The provided error messages are *insufficient*, and EqFix could not generate a useful equation pattern.
- Restricted DSL expressiveness (4 cases): The test case is *deviated far* from the training examples of the same group, and due to the restricted expressiveness of our DSL, the learned rule could not generalize to that test case.

Testing the Rule Library Additionally, we conducted another evaluation on the same dataset, but only for EqFix, that fitted a more realistic setting where equation repair problems are solved by trying the initial rules saved in a rule library. We obtained the initial rule library by learning from the entire training set examples (C4) under the training mode. Then, under the applying mode, we tested all 89 test cases. Interestingly, compared with the results shown in Table 3, one more test case (#71) was solved (by the top-ranked rule synthesized from the example group #69).

6 Related Work

Program Repair Automated program repair aims to automatically correct programs so that they satisfy the desired specification [5]. Heuristic-based repair tools such as GenProg [26,4,15] employs an extended form of genetic programming with heuristics. However, it is shown that these techniques produce patches that overfit the test suite [25]. Ranking techniques have been studied to address the problem. ACS [27] produces precise patches with a refined ranking technique for condition synthesis. PAR [13] mines bug fix patterns from the history and gives frequently occurring fixes high priority. Prophet [17] outperforms the previous works by learning a probabilistic model for ranking the candidate patches.

Semantics-based repair techniques generate repairs via symbolic execution [21,18] and program synthesis [16]. Such techniques, however, are also suspected of overfitting the test suite. Recently, a new repair synthesis engine called S3 is proposed [14]. It leverages *Programming by Examples* (PBE) methodology to synthesize high-quality bug repairs, elaborating several ranking features.

Technically, EqFix belongs to the semantics-based family. To avoid overfitting, we also rely on the ranking technique for promoting rules with a high generality. Furthermore, syntactic errors are common and must be tackled in equation repair, while it is usually neglected in program repair, as people concern more about bugs [19]. HelpMeOut [9] aids developers to debug compilation

and run-time error messages by suggesting past solutions. Unlike EqFix, it only provides related examples and cannot create repairs for new problems.

NoFAQ [2] is a system for fixing buggy Unix commands using PBE. Although the problem domain is similar to ours, due to its lazy synthesis algorithm, it only can synthesize a practical fixing rule when at least two examples are provided. In contrast, in many cases, one example is sufficient for EqFix. The DSL of NoFAQ can only accept tokenized strings (separated by spaces) as input. However, there is no direct way to tokenize an erroneous equation in our problem domain. We thus introduce the equation patterns and propose a mechanism for synthesizing them. These equation patterns help to pattern match against an equation and extract the variant parts, which need to be transformed later.

Text Transformation FlashFill [7] pioneered in text transformation via program synthesis and was later extended for semantic transformation [24]. A similar technique is put into a live programming environment by StriSynth [6]. FlashExtract [12] automates data extraction by highlighting texts on web pages. String transformation is performed at a high level in these techniques, but it is unsuitable for repairing equations. We realize that error messages guide the repairing process. By pattern matching the error messages, we only perform the transformation on a few variables instead of the entire equation, which takes less time.

VSA-based Program Synthesis Version space algebra (VSA) has been widely adopted in PBE applications [11,10,22,1,28,23]. In those applications, VSA is critical as the set of candidate programs is possibly very large. In EqFix, the synthesized rules are represented compactly with VSA.

VSA-based program synthesis has also been applied in program transformation. Refazer [23] is a framework that automatically learns program transformations at an abstract syntax tree level. Feser et al. [3] propose a method for example-guided synthesis of recursive data structure transformations in functional programming languages. Nguyen et al. [20] present a graph-based technique that guides developers in adapting API usages.

7 Conclusion & Future Directions

We present EqFix, a system for fixing both compilation and typesetting errors in \LaTeX equations. We design a DSL for expressing fixing rules and propose a synthesis algorithm to learn rules from user-provided examples. In the future, our tool can be improved by leveraging data from various sources like \LaTeX online forums via crowdsourced learning. When a large rule library is established, it would be interesting to develop an EqFix plugin in modern editors for practical use. Furthermore, because adding more data does not require any change in EqFix but simply needs more input-output examples to construct synthesis rules, our approach can potentially be applied to other string and mathematical equation systems.

References

1. Barman, S., Chasins, S., Bodik, R., Gulwani, S.: Ringer: web automation by demonstration. In: Proceedings of the 2016 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications. pp. 748–764. ACM (2016)
2. D’Antoni, L., Singh, R., Vaughn, M.: NoFAQ: Synthesizing command repairs from examples. In: Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. pp. 582–592. ESEC/FSE 2017, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3106237.3106241>
3. Feser, J.K., Chaudhuri, S., Dillig, I.: Synthesizing data structure transformations from input-output examples. In: Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation. pp. 229–239. PLDI ’15, ACM, New York, NY, USA (2015). <https://doi.org/10.1145/2737924.2737977>
4. Goues, C.L., Nguyen, T., Forrest, S., Weimer, W.: GenProg: A generic method for automatic software repair. *IEEE Transactions on Software Engineering* **38**(1), 54–72 (Jan 2012). <https://doi.org/10.1109/TSE.2011.104>
5. Goues, C., Forrest, S., Weimer, W.: Current challenges in automatic software repair. *Software Quality Journal* **21**(3), 421–443 (Sep 2013). <https://doi.org/10.1007/s11219-013-9208-0>
6. Gulwani, S., Mayer, M., Niksic, F., Piskac, R.: StriSynth: Synthesis for live programming. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering. vol. 2, pp. 701–704 (May 2015). <https://doi.org/10.1109/ICSE.2015.227>
7. Gulwani, S.: Automating string processing in spreadsheets using input-output examples. In: Proceedings of the 38th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. pp. 317–330. POPL ’11, ACM (2011). <https://doi.org/10.1145/1926385.1926423>
8. Gulwani, S., Esparza, J., Grumberg, O., Sickert, S.: Programming by examples (and its applications in data wrangling). *Verification and Synthesis of Correct and Secure Systems* (2016)
9. Hartmann, B., MacDougall, D., Brandt, J., Klemmer, S.R.: What would other programmers do: Suggesting solutions to error messages. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1019–1028. CHI ’10, ACM, New York, NY, USA (2010). <https://doi.org/10.1145/1753326.1753478>
10. Kini, D., Gulwani, S.: FlashNormalize: Programming by examples for text normalization. In: Proceedings of the 24th International Conference on Artificial Intelligence. pp. 776–783. AAAI Press (2015)
11. Lau, T.A., Domingos, P., Weld, D.S.: Version space algebra and its application to programming by demonstration. In: Proceedings of the Seventeenth International Conference on Machine Learning. pp. 527–534. ICML ’00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2000)
12. Le, V., Gulwani, S.: FlashExtract: A framework for data extraction by examples. In: Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation. pp. 542–553. PLDI ’14, ACM (2014). <https://doi.org/10.1145/2594291.2594333>
13. Le, X.B.D., Lo, D., Goues, C.L.: History driven program repair. In: 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER). vol. 1, pp. 213–224 (March 2016). <https://doi.org/10.1109/SANER.2016.76>

14. Le, X.B.D., Chu, D.H., Lo, D., Le Goues, C., Visser, W.: S3: Syntax- and semantic-guided repair synthesis via programming by examples. In: Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. pp. 593–604. ESEC/FSE 2017, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3106237.3106309>
15. Le Goues, C., Dewey-Vogt, M., Forrest, S., Weimer, W.: A systematic study of automated program repair: Fixing 55 out of 105 bugs for \$8 each. In: Proceedings of the 34th International Conference on Software Engineering. pp. 3–13. ICSE '12, IEEE Press, Piscataway, NJ, USA (2012)
16. Long, F., Rinard, M.: Staged program repair with condition synthesis. In: Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. pp. 166–178. ESEC/FSE 2015, ACM, New York, NY, USA (2015). <https://doi.org/10.1145/2786805.2786811>
17. Long, F., Rinard, M.: Automatic patch generation by learning correct code. In: Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. pp. 298–312. POPL '16, ACM, New York, NY, USA (2016). <https://doi.org/10.1145/2837614.2837617>
18. Mehtaev, S., Yi, J., Roychoudhury, A.: Angelix: Scalable multiline program patch synthesis via symbolic analysis. In: Proceedings of the 38th International Conference on Software Engineering. pp. 691–701. ICSE '16, ACM, New York, NY, USA (2016). <https://doi.org/10.1145/2884781.2884807>
19. Monperrus, M.: Automatic software repair: A bibliography. *ACM Comput. Surv.* **51**(1), 17:1–17:24 (Jan 2018). <https://doi.org/10.1145/3105906>
20. Nguyen, H.A., Nguyen, T.T., Wilson, Jr., G., Nguyen, A.T., Kim, M., Nguyen, T.N.: A graph-based approach to API usage adaptation. In: Proceedings of the ACM International Conference on Object Oriented Programming Systems Languages and Applications. pp. 302–321. OOPSLA '10, ACM, New York, NY, USA (2010). <https://doi.org/10.1145/1869459.1869486>
21. Nguyen, H.D.T., Qi, D., Roychoudhury, A., Chandra, S.: SemFix: Program repair via semantic analysis. In: Proceedings of the 2013 International Conference on Software Engineering. pp. 772–781. ICSE '13, IEEE Press, Piscataway, NJ, USA (2013)
22. Polozov, O., Gulwani, S.: FlashMeta: A framework for inductive program synthesis. *ACM SIGPLAN Notices* **50**(10), 107–126 (2015)
23. Rolim, R., Soares, G., D'Antoni, L., Polozov, O., Gulwani, S., Gheyi, R., Suzuki, R., Hartmann, B.: Learning syntactic program transformations from examples. In: Proceedings of the 39th International Conference on Software Engineering. pp. 404–415. ICSE '17, IEEE Press, Piscataway, NJ, USA (2017). <https://doi.org/10.1109/ICSE.2017.44>
24. Singh, R., Gulwani, S.: Learning semantic string transformations from examples. *Proc. VLDB Endow.* **5**(8), 740–751 (Apr 2012). <https://doi.org/10.14778/2212351.2212356>
25. Smith, E.K., Barr, E.T., Le Goues, C., Brun, Y.: Is the cure worse than the disease? overfitting in automated program repair. In: Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. pp. 532–543. ESEC/FSE 2015, ACM, New York, NY, USA (2015). <https://doi.org/10.1145/2786805.2786825>
26. Weimer, W., Nguyen, T., Goues, C.L., Forrest, S.: Automatically finding patches using genetic programming. In: 2009 IEEE 31st International Conference on Software Engineering. pp. 364–374 (May 2009). <https://doi.org/10.1109/ICSE.2009.5070536>

27. Xiong, Y., Wang, J., Yan, R., Zhang, J., Han, S., Huang, G., Zhang, L.: Precise condition synthesis for program repair. In: Proceedings of the 39th International Conference on Software Engineering. pp. 416–426. ICSE '17, IEEE Press, Piscataway, NJ, USA (2017). <https://doi.org/10.1109/ICSE.2017.45>
28. Yaghmazadeh, N., Klinger, C., Dillig, I., Chaudhuri, S.: Synthesizing transformations on hierarchically structured data. In: Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation. pp. 508–521. PLDI '16, ACM (2016). <https://doi.org/10.1145/2908080.2908088>
29. Zhu, F., He, F.: Eqfix: Fixing latex equation errors by examples (2022). <https://doi.org/10.48550/ARXIV.2107.00613>, <https://arxiv.org/abs/2107.00613>