Формализация математических текстов через теорию дискурса

Хасанянова Сания

Научный руководитель: Перепечко Александр Юрьевич

Были поставлены следующие задачи:

1. Изучить теорию дискурса; разобраться в понятиях теории дискурса.

2. Попытаться расширить теорию дискурса на математический язык.

Теория дискурса (или DRT)

Теория дискурса – это логика, используемая для описания (семантического содержания) естественного языка. Её элементарной единицей является структура репрезентации дискурса (или DRS), например, следующая:

$x \ y$
$John(x)$ $dog(y)$ $owns(x, y)$

John owns a dog.

$$\exists x.\exists y.(John(x) \wedge dog(y) \wedge owns(x, y))$$

Анафора (лингвистика)

Анафора – отношение между языковыми выражениями, при котором в смысл одного выражения входит отсылка к другому, ранее упомянутому языковому выражению.

Первый член анафорического отношения называется антецедентом, второй – анафо́ром. Высказывание, включающее анафор без антецедента, даже синтаксически законченное, обладает смысловой неполнотой.

Теория дискурса (или DRT)

Теория дискурса – это логика, используемая для описания (семантического содержания) естественного языка. Её элементарной единицей является структура репрезентации дискурса (или DRS), например, следующая:

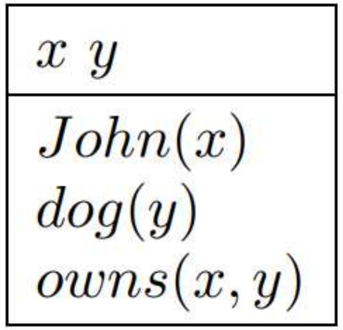
$x \ y$
$John(x)$ $dog(y)$ $owns(x, y)$

John owns a dog.

$$\exists x. \exists y. (John(x) \wedge dog(y) \wedge owns(x, y))$$

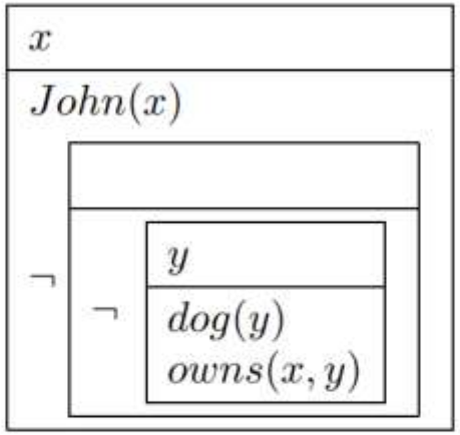
John owns a dog. **He** used to own a cat.

John owns a dog. **It** is called Fido.



John owns a dog.

$\exists x.\exists y.(John(x) \wedge dog(y) \wedge owns(x, y))$



As for John, it is not the case that he does not own a dog.

$\exists x.(John(x) \wedge \neg\neg\exists y.(dog(y) \wedge owns(x, y)))$

As for John, it is not the case that he does not own a dog. **He** used to own a cat.

As for John, it is not the case that he does not own a dog. **It** is called Fido.

DRS Δ

$$
\begin{array}{|l|}
\hline
x_1 \ \ldots \ x_n \\
\hline
\gamma_1 \\
\vdots \\
\gamma_m \\
\hline
\end{array}
$$

x_i − референты DRS, а каждое условие DRS γ_i является одним из следующих:

1. Атомарный предикат, оцениваемый на референтах, которые доступны из Δ. (Доступность определяется чуть позже).

2. Уравнение x = x’, где x и x’ − референты, доступные из Δ.

3. ¬Γ, где Γ − DRS.

4. Γ ⇒ Γ’, где Γ и Γ’ − DRS.

Доступность

DRS может получить доступ к DRS, которые содержат его, а также к DRS, которые влекут его за собой.

Говорим, что референт x доступен из DRS Γ тогда и только тогда, когда DRS Δ, в которую вводится x, доступна из Γ.

Обновление дискурса

John owns a dog. Mary owns a cat.

$$
\begin{array}{|l|}
\hline
\quad \\
\hline
\quad \\
\quad \\
\hline
\end{array}
\qquad
\begin{array}{|l|}
\hline
x \ y \\
\hline
John(x) \\
dog(y) \\
owns(x, y) \\
\hline
\end{array}
\qquad
\begin{array}{|l|}
\hline
z \ u \\
\hline
Mary(z) \\
cat(u) \\
owns(z, u) \\
\hline
\end{array}
\qquad
\begin{array}{|l|}
\hline
x \ y \ z \ u \\
\hline
John(x) \\
dog(y) \\
owns(x, y) \\
Mary(z) \\
cat(u) \\
owns(z, u) \\
\hline
\end{array}
$$

John owns a dog. He is hairy.

$x\ y$
$John(x)$
$dog(y)$
$owns(x, y)$

$\underline{u}$
$hairy(\underline{u})$

$x\ y\ \underline{u}$
$John(x)$
$dog(y)$
$owns(x, y)$
$hairy(\underline{u})$

$x\ y\ \underline{u}$
$John(x)$
$dog(y)$
$owns(x, y)$
$hairy(\underline{u})$
$\underline{u}$=y

$$\begin{array}{|l|}\hline x_1 \ \ \dots \ \ x_n \\ \hline \gamma_1 \\ \vdots \\ \gamma_m \\ \hline \end{array} \ \uplus \ \begin{array}{|l|}\hline x'_1 \ \ \dots \ \ x'_{n'} \\ \hline \gamma'_1 \\ \vdots \\ \gamma'_{m'} \\ \hline \end{array} \ = \ \begin{array}{|l|}\hline x_1 \ \ \dots \ \ x_n \ x'_1 \ \ \dots \ \ x'_{n'} \\ \hline \gamma_1 \\ \vdots \\ \gamma_m \\ \gamma'_1 \\ \vdots \\ \gamma'_{m'} \\ \hline \end{array}$$

$$D_n = D_{n-1} \uplus S_n$$

Следующая задача – попытка расширить теорию дискурса на математический язык

3 + 4 is prime.

$x \; y \; z$
$3(x)$
$4(y)$
$sum(z, x, y)$
$prime(z)$

Проблема: в таком случае 3 и 4 доступны в качестве антецедентов для анафоры.

Решение

x
$x = sum(3(), 4())$ $prime(x)$

Это представление правильно предсказывает, что 3 + 4 является доступным антецедентом для анафоры, а 3 или 4 − нет.

Представление переменных

If x and y are real numbers, then $x^2 - y^2 = (x + y)(x - y)$.

Пара предложений, которые имеют одинаковый смысл:

If n is a natural number which is greater than 1, then there is a prime p such that $p|n$.

and

Every natural number which is greater than 1 has a prime divisor.

$$\begin{array}{|c|}\hline \\ \hline \begin{array}{|l|}\hline x\ y \\ \hline \mathit{natural\_number}(x) \\ y = \mathit{1}() \\ \mathit{greater\_than}(x, y) \\ \hline\end{array} \Rightarrow \begin{array}{|l|}\hline z \\ \hline \mathit{is\_prime}(z) \\ \mathit{divides}(z, x) \\ \hline\end{array} \\ \hline\end{array}$$

Every natural number which is greater than 1 has a prime divisor.

$$\begin{array}{|c|}\hline \\ \hline \begin{array}{|l|}\hline \mathbf{n}\ y \\ \hline \mathit{natural\_number}(\mathbf{n}) \\ y = \mathit{1}() \\ \mathit{greater\_than}(\mathbf{n}, y) \\ \hline\end{array} \Rightarrow \begin{array}{|l|}\hline \mathbf{p} \\ \hline \mathit{is\_prime}(\mathbf{p}) \\ \mathit{divides}(\mathbf{p}, \mathbf{n}) \\ \hline\end{array} \\ \hline\end{array}$$

If n is a natural number which is greater than 1, then there is a prime p such that $p|n$.

Дальнейшая работа

Сравнительный анализ существующих подходов представления математического языка.

Зачем это всё надо?

Спасибо за внимание!