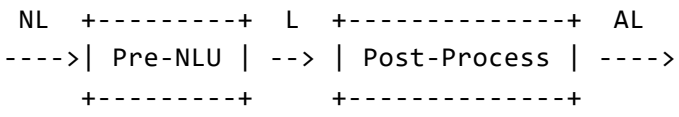
Annotation Handbook v2

This is a handbook for annotation. Contents are mostly collected from weekly meetings since the project starts. We modified some minor parts in the 2nd version.

Overview

Pipeline



- NL: Natural Language
- L: Middle Language
- AL: Assertional Logic

The annotation task is to convert NL to L manually.

Principle

The annotation should achieve:

1. **No ambiguity.** With the information inside the annotations, we can work out the solution by hand.
2. **Apply basic AL syntax.** Conversion to AL should be possible and easy.
3. **Close to NL.** It should represent the question without rephrasing it.

Intro to AL

AL is short for Assertional Logic (https://linkspringer.53yu.com/chapter/10.1007/978-3-319-63703-7_9). An old version of the paper is here (https://arxiv.org/abs/1701.03322). We mostly use a subset syntax of AL as the syntax of our annotation.

The basic structure of AL domain is composed of **Individual**, **Concept** and **Operator**. Individuals represent objects in the domain, concepts represent groups of objects sharing something in common, while operators represents relationships and connections among individuals and concepts. Concepts are sets of individuals, here we usually use concepts to declare variables. Operators are like functions (but actually more powerful than those in first order logic).

AL is human-friendly and easy-to-read. See some examples and you'll understand how it works.

Annotation Structure

An annotation is composed of 4 parts:

1. NL. The natural language representation of the question;
2. Fact List. A list of assertions representing the question.
3. Query List. A list of terms representing the queries.
4. Answer. A list of terms representing the answer.
5. Spans. The span in natural language corresponding to each translated logic expression (assertion).

For some questions, the annotation may not exist. See the last part 'Cannot Annotate' for details.

Natural Language

The questions in natural language are composed of 2 parts: Chinese language text and LaTeX math expressions. We require that:

1. The Chinese texts are consistent with those in the images.
2. The Chinese texts are clean. No misspelling.
3. All math expressions (include the numbers) are written in LaTeX. They have to be bracketed in dollar signs ($).
4. No Chinese characters are bracketed in dollar signs.
5. The question should use the question mark (?, English symbol) as the unknown part instead of the underlines.
6. Extra spaces does not matter.

Some natural language questions might be noisy. You should clean up the natural language first, then continue the annotation.

Original Text:

已知: M, N 两点关于 y轴对称, 点 M 的坐标为 (a, b), 且点 M 在双曲线 $y=\frac{1}{x}$ 上, 点N 在直线 $\mathrm{y}=\mathrm{x}+3$ 上,MN长为3。设则抛物线 $\mathrm{y}=-abx^2+(a+b)x$ 的顶点坐标是.

```
已知: $\mathrm{M}, \mathrm{N}$ 两点关于 y轴对称, 点 $\mathrm{M}$ 的坐标为 $(\mathrm{a}, \mathrm{b})$, 且点 M 在双曲线 $y=\frac{1}{x}$ 上, $点 \mathrm{N}$ 在直线 $\mathrm{y}=\mathrm{x}+3$ 上,$MN$长为3。设则抛物线 $\mathrm{y}=$ $-a b x^{2}+(a+b) x$ 的顶点坐标是.
```

Cleaned up:

已知: M, N 两点关于 y轴对称, 点 M 的坐标为 (a, b), 且点 M 在双曲线 $y=\frac{1}{x}$ 上, 点N 在直线 $y=x+3$ 上,MN长为3。设则抛物线 $y=-abx^2+(a+b)x$ 的顶点坐标是?

```
已知: $M, N$ 两点关于 $y$轴对称, 点 $M$ 的坐标为 $(a,b)$, 且点 $M$ 在双曲线 $y=\frac{1}{x}$ 上, 点$N$ 在直线 $y=x+3$ 上,$MN$长为$3$。设则抛物线 $y=-a b x^{2}+(a+b) x$ 的顶点坐标是?
```

Original Text:

设椭圆M:$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1(a>b>0)$右顶点和上顶点分别为$A_1, A_2$

```
设椭圆M:$\frac { x ^ { 2 } } { a ^ { 2 } } + \frac { y ^ { 2 } } { b ^ { 2 } } = 1 ( a > b > 0 )$右顶点和上顶点分别为$A_1,A_2$
```

Cleaned up:

设椭圆M:$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1(a>b>0)$右顶点和上顶点分别为$A_1$、$A_2$

```
设椭圆$M$:$\frac { x ^ { 2 } } { a ^ { 2 } } + \frac { y ^ { 2 } } { b ^ { 2 } } = 1 ( a > b > 0 )$右顶点和上顶点分别为$A_1$、$A_2$
```

What's more, sometimes we need to fix errors or noise in the text.

1. Remove the serial number. (e.g. `1. 椭圆... -> 椭圆...`)
2. If a symbol should be subscript (e.g. F_1 `$F_1$`) or superscript (e.g. n^2 `$n^2$`) but the text is flat (e.g. $F1$ `$F1$`, $n2$ `$n2$`), it is required to fix it.
3. Some symbols might get the dollar signs ($) missing (e.g. `椭圆M`). Add them back (e.g. `椭圆$M$`).
4. Split variable declarations into seperate dollar sign spans. (e.g. `$A, B$ -> $A$、$B$`)
5. If possible, fix some special LaTeX symbols (`\mathrm`, `\mid`, etc.)

Example:

Original Text:

1.已知双曲线的焦点在 x 轴上，坐标为$(0，\frac{5}{2})$，且 $a+c=9$，$b=3$，则它的标准方程是

```
1.已知双曲线的焦点在 $x$ 轴上，坐标为$(0，\frac{5}{2})$，且 $a+c=9 ， b=3$ ，则它的标准方程是
```

Cleaned up:

已知双曲线的焦点在 x 轴上，坐标为$(0, \frac{5}{2})$，且 $a+c=9$ ， $b=3$ ，则它的标准方程是?

```
已知双曲线的焦点在 $x$ 轴上，坐标为$(0,\frac{5}{2})$，且 $a+c=9$ ， $b=3$ ，则它的标准方程是?
```

Example:

Original Text:

已知双曲线的两个焦点 $F1(-\sqrt{10}，0)$，$F2(\sqrt{10}，0)$，P 是此双曲线上的一点，且 $\overrightarrow{PF_1} \cdot \overrightarrow{PF_2}=0$ ，$|PF1| \cdot |PF_2 |=2$ ，则该双曲线的方程是

```
已知双曲线的两个焦点 $F{1}(-\sqrt{10} ， 0) ， F2( \sqrt{10} ， 0) ， P$ 是此双曲线上的一点，且$\overrightarrow{PF_1} \cdot \overrightarrow{PF_2}=0$ ，$| PF1| \cdot| PF_2 \mid=2$ ，则该双曲线的方程是
```

Cleaned up:

已知双曲线的两个焦点 $F_1(-\sqrt{10}, 0)$ 、 $F_2(\sqrt{10}, 0)$ ， P 是此双曲线上的一点，且 $\overrightarrow{PF_1} \cdot \overrightarrow{PF_2}=0$ ，$|PF_1| \cdot |PF_2| = 2$ ，则该双曲线的方程是?

```
已知双曲线的两个焦点 $F_{1}$$(-\sqrt{10}, 0)$ 、 $F_2$$( \sqrt{10}, 0)$ ， $P$ 是此双曲线上的一点，且$\overrightarrow{PF_1} \cdot \overrightarrow{PF_2}=0$ ，$| PF_1| \cdot| PF_2 |=2$ ，则该双曲线的方程是?
```

Example:

Original Text:

双曲线的焦点在x轴上，实轴长为6，虚轴长为8，则双曲线的标准方程是_____

```
双曲线的焦点在$\mathrm{x}$轴上，实轴长为6，虚轴长为8，则双曲线的标准方程是_____
```

Cleaned up:

双曲线的焦点在x轴上，实轴长为6，虚轴长为8，则双曲线的标准方程是?

```
双曲线的焦点在$x$轴上，实轴长为$6$，虚轴长为$8$，则双曲线的标准方程是?
```

We have added scripts to automatically fix most of the problems. But annotators still need to check whether the text is consistent with the requirements above.

Attention: You MUST clean up the natural language before selecting spans for each annotated sentence.

Syntax

The syntax of our annotation language:

Basic Syntax

```
Sentence    -> Assertion
Assertion   -> Term = Term
Term        -> Operator(Terms) | AtomicIndividual | (Assertion) | (Terms) | {Terms}
Terms       -> Term | Terms, Term

AtomicIndividual  -> Constant | Variable
Constant    -> 1 | 2 | True | False | pi | e ...
Variable    -> Parabola_C | Point_A ...

Operator    -> In | PointOnCurve
             | Radius | Length | Sin
             | Focus | Apex | ...
```

Variable Declaration

This should be clear. Variables declare in this way:

```
var[, vars...]: Concept
```

Syntactic Sugar

We use syntactic sugar (without ambiguity) in the annotation. This includes

Symbol	Code	Comments
$=$	=	
$<$	<	
$>$	>	
$\leq$	<=	
$\geq$	>=	
$+$	+	
$-$	-	
$\times$	*	
$\div$	/	
a^b	**,^	power
$\neq$		Not allowed! use `Negation(A=B)` instead.
$\wedge$	&	Same as `And(A, B, ...)`

You are allowed to drop `= True` for predicates.

Example:

With `= True`:

```
(a > 0) = True
IsParallel(l1, l2) = True
```

Without `= True`:

```
a > 0
IsParallel(l1, l2)
```

We usually drop `= True` for inequality syntactic sugar and keep it for other situations. But it doesn't matter, actually.

Some Tips

1. In the fact list, a sentence is either an assertion (`... = ...`) or a declaration (`... : ...`).

2. In the annotation system, you do NOT need to write `= ?` in the query list. Each line should be a term instead of assertion.
3. The annotation is not sensitive in the order. It doesn't matter which translated sentence comes first, so do the declarations.
4. Variable names doesn't matter, but we recommend to use the same variable names as those in the questions if possible. Notice that variable naming only allows letters (`a-zA-Z`), numbers (`0-9`) or underscore (`_`) and must starts with a letter.

Example:

椭圆C

✔:

```
C: Parabola
```

✔:

```
C : Parabola
```

✔:

```
C :Parabola
```

椭圆C_1

✔:

```
C_1: Parabola
```

✔:

```
C1: Parabola
```

椭圆C'

✔:

```
C1: Parabola
```

✔:

```
C_: Parabola
```

✘:

```
C': Parabola
```

Individual, Concept, Operator Lookup Table

Individual

Name	Description
axis	坐标轴
xAxis	x轴
yAxis	y轴
oo	infinity
rad	弧度
degree	度
pi	3.14 π
pm	$\pm$

Concept

Name	Description
Angle	角
Real	实数
Number	数
Origin	原点
Vector	向量
Curve	曲线
Triangle	三角形
Axis	坐标轴
Ray	射线
LineSegment	线段
Circle	圆
Parabola	抛物线
Hyperbola	双曲线
Ellipse	椭圆
ConicSection	圆锥曲线
Line	直线
Point	点

Operator

You may look up the operators on the annotation website (http://47.102.141.251/#/docs).

Span

The mapping from translated language to natural language is useful in model training. As a result, for each assertion in translated language, span(s) is also required. A span is a mapping from natural language to assertion. It consists of a minimal part of natural language that can be translated to the assertion. For one assertion there may exist multiple spans, especially for declarations.

a. Variable Declaration

The span corresponding to a variable declaration is all the mentions in the question text that represent this variable. Notice that we treat the math expression between dollar signs ($) as a group that cannot be devided.

Example:

椭圆$\frac{x^2}{a^2}+\frac{y^2}{16}=1$的一个焦点为$(3,0)$,则这个椭圆的方程为?

```
C: Parabola
    [(0,64), (82,84)]  ([[椭圆$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } {
a: Number
    [(2, 64)] ([[$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1$]])
```

We want the span to contain:

1. Chinese naming (must be consistent with the concept);
2. Variable representation;
3. Expression / Coordinate.

Example:

椭圆$\frac{x^2}{a^2}+\frac{y^2}{16}=1$

```
椭圆$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1$
```

```
C: Parabola
    [(0,64)]  ([[椭圆$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1
```

椭圆C:$\frac{x^2}{a^2}+\frac{y^2}{16}=1$

```
椭圆$C$:$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1$
```

```
C: Parabola
    [(0,68)]  ([[椭圆$C$:$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 }
```

椭圆$C:\frac{x^2}{a^2}+\frac{y^2}{16}=1$

```
椭圆$C:\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1$
```

```
C: Parabola
    [(0,66)]  ([[椭圆$C:\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } =
```

焦点在x轴上的椭圆C的离心率为$\frac{1}{2}$

```
焦点在$x$轴上的椭圆$C$的离心率为$\frac{1}{2}$
```

```
C: Parabola
    [(9,14)]  ([[椭圆$C$]])
```

两曲线相交于点$(2,2)$。

```
两曲线相交于点$(2,2)$。
```

```
P: Point
    [(6,14)]  ([[点$(2,2)$]])
```

两曲线相交于点$P\ (2,2)$。

```
两曲线相交于点$P$ $(2,2)$。
```

```
P: Point
    [(6,18)]  ([[点$P$ $(2,2)$]])
```

两曲线相交于坐标原点O。

两曲线相交于坐标原点O。

```
O: Origin
    [(6,14)]  ([[坐标原点$O$]])
```

But they must be consecutive:

Example:

坐标原点O

坐标原点O

```
O: Origin
    [(0,7)]  ([[坐标原点$O$]])
```

O为坐标原点

O为坐标原点

```
O: Origin
    [(0,3)]  ([[$O$]])
```

Example:

点P的坐标为$(2, 2)$

点P的坐标为$(2,2)$

✔:

```
P: Point
    [(0,4)]  ([[点$P$]])
```

✘:

```
P: Point
    [(0,15)]  ([[点$P$的坐标为$(2,2)$]])
```

Example:

若$\frac{x^2}{1+m}+\frac{y^2}{1-m}=1$表示双曲线，则$m$的取值范围是?

若$\frac{x^{2}}{1+m}+\frac{y^{2}}{1-m}=1$表示双曲线，则$m$的取值范围是?

✔:

```
E: Hyperbola
    [(42,45)]  ([[双曲线]])
Expression(E) = (x**2/(m + 1) + y**2/(1 - m) = 1)
    [(1,45)]  ([[$\frac{x^{2}}{1+m}+\frac{y^{2}}{1-m}=1$表示双曲线]])
m: Number
    [(47,50)] ([[$m$]])
```

✘:

```
E: Hyperbola
    [(1,45)]  ([[$\frac{x^{2}}{1+m}+\frac{y^{2}}{1-m}=1$表示双曲线]])
```

✘:

```
m: Number
    [(1,40), (47,50)] ([[$\frac{x^{2}}{1+m}+\frac{y^{2}}{1-m}=1$], [$m$]])
```

Besides, notice that m shows up as a single token in the sentence. At this time, we take this token as m's representation, ignoring the expressions contains m.

Sometimes the Chinese naming is differnt from the concept name. We do NOT contain the Chinese naming under this circumstance.

Example:

抛物线C的焦点$P(0,2)$在抛物线E上

抛物线C的焦点$P$$(0,2)$在抛物线$E$上

✔:

```
P: Point
    [(9,19)]  ([[$P$$(0,2)$]])
```

✘:

```
P: Point
    [(7,19)]  ([[焦点$P$$(0,2)$]])
```

✘:

```
P: Point
    [(8,19)]  ([[点$P$$(0,2)$]])
```

✘:

```
P: Point
    [(9,12)]  ([[$P$]])
```

Notice that `焦点` is one single word and it does not represent the concept of `Point`.

If two variables show up at the same time, the span depends on the constituent structure of the sentence.

Example:

已知双曲线的两个焦点$F_1(-\sqrt{10},0)$，$F_2(\sqrt{10},0)$

```
已知双曲线的两个焦点$F_1$$(-\sqrt{10}, 0)$，$F_2$$(\sqrt{10}, 0)$
```

✔:

```
F1: Point
    [(10,32)]  ([[$F_1$$(-\sqrt{10}, 0)$]])
F2: Point
    [(33,54)]  ([[$F_2$$(\sqrt{10}, 0)$]])
```

✘:

```
F1: Point
    [(8,32)]  ([[焦点$F_1$$(-\sqrt{10}, 0)$]])
```

✘:

```
F1: Point
    [(8,15)]  ([[焦点$F_1$]])
```

✘:

```
F1, F2: Point
    [(6,54)]  ([[两个焦点$F_1$$(-\sqrt{10}, 0)$，$F_2$$(\sqrt{10}, 0)$]])
```

Here, `两个焦点` modifies `$F_1$`, `$F_2$` at the same time. So `两个焦点` should neither be covered by the span of `F1` nor by `F2`. Also, `焦点` does not represent the concept of `Point`.

Example:

已知两个点F_1、F_2

已知两个点F_1、F_2

✔:

```
F1: Point
    [(5,10)]  ([[$F_1$]])
F2: Point
    [(11,16)]  ([[$F_2$]])
```

✘:

```
F1: Point
    [(4,10)]  ([[点$F_1$]])
```

✘:

```
F1, F2: Point
    [(2,16)]  ([[两个点$F_1$、$F_2$]])
```

Here, `两个点` modifies `$F_1$`, `$F_2$` at the same time. So `两个点` should neither be covered by the span of `F1` nor by `F2`.

Example:

过双曲线C的左焦点F_1且斜率为$\frac{1}{3}$的直线l交双曲线C的左右两支于A、B两点

过双曲线C的左焦点F_{1}且斜率为$\frac{1}{3}$的直线l交双曲线C的左右两支于A、B两点

✔:

```
A: Point
    [(54,57)]  ([[$A$]])
B: Point
    [(58,61)]  ([[$B$]])
```

✘:

```
B: Point
    [(58,63)]  ([[$B$两点]])
```

When we declare numbers (e.g. `a`, `e`), we first find tokens in the sentence. If none exists, choose the expression that contains the number.

Example:

椭圆$\frac{x^2}{a^2}+\frac{y^2}{16}=1$的离心率为$\frac{1}{2}$，实数$a$的值为?

```
a: Real
    [(83, 88)] ([[实数$a$]])
```

椭圆$\frac{x^2}{a^2}+\frac{y^2}{16}=1$的离心率为$\frac{1}{2}$，椭圆的准线方程为?

```
a: Number
    [(2, 64)] ([[$\frac { x ^ { 2 } } { a^2 } + \frac { y ^ { 2 } } { 16 } = 1$]])
```

b. Assertions

A span for an assertion is the minimal part in natural language that can be translated to this assertion. In most cases, an assertion has only one corresponding span in the question text.

Remember to annotate both facts and queries. If possible, do not cover the stop words in the spans.

Example:

双曲线C与椭圆$\frac{x^2}{36}+\frac{y^2}{16}=1$有相同的焦点,且$C$的渐近线为$x\pm\sqrt{3}y=0$,则双曲线$C$的方程?

✔:

```
- facts:
C: Hyperbola
    [(0,6), (78,81), (108,114)]  ([[双曲线$C$], [C], [双曲线$C$]])
E: Ellipse
    [(7,9)]  ([[椭圆]])
Expression(E) = (x**2/36 + y**2/16 = (1))
    [(7, 70)]  ([[椭圆$\frac { x ^ { 2 } } { 36 } + \frac { y ^ { 2 } } { 16 } = 1
Focus(C) = Focus(E)
    [(0,76)]  ([[双曲线$C$与椭圆$\frac { x ^ { 2 } } { 36 } + \frac { y ^ { 2 } } {
Expression(Asymptote(C)) = {x+sqrt(3)*y=0, x-sqrt(3)*y=0}
    [(78,106)]  ([[$C$的渐近线为$x\pm \sqrt{3}y = 0$]])
```

✘:

```
Focus(C) = Focus(E)
    [(0,9) (70,76)]  ([[双曲线$C$与椭圆], [有相同的焦点]])
```

If there are pronouns (它) or mentions, assume that they carry the information of the corresponding entities.

若双曲线的渐近线方程为$y=\pm 3x$,它的一个焦点是$(\sqrt{10},0)$,则双曲线的标准方程是?

✔:

```
- facts:
C_1 : Hyperbola
    [(1,4), (21,22), (45,48)]  ([双曲线, 它, 双曲线])
Expression(Asymptote(C_1)) = (y = pm*3*x)
    [(1,20)] ([双曲线的渐近线方程为$y=\pm3x$])
F : Point
    [(23,27)] ([一个焦点])
Coordinate(F) = (sqrt(10),0)
    [(21,43)] ([它的一个焦点是$(\sqrt{10},0)$])
In(F,Focus(C_1)) = True
    [(21,27)] ([它的一个焦点])
```

✘:

```
Coordinate(F) = (sqrt(10),0)
    [(1,43)] ([双曲线的渐近线方程为$y=\pm3x$,它的一个焦点是$(\sqrt{10},0)$])
```

If possible, we will ignore the modifiers.

Example:

过点 F 且倾斜角为 $\frac{\pi}{6}$ 的直线 l 与抛物线 C 交于第一象限点 A

```
过点 $F$ 且倾斜角为 $\frac{\pi}{6}$ 的直线 $l$ 与抛物线 $C$ 交于第一象限点 $A$
```

✔:

```
Intersection(l, C) = A
    [(30,57)] ([直线 $l$ 与抛物线 $C$ 交于第一象限点 $A$])
```

✘:

```
Intersection(l, C) = A
    [(0,57)] ([过点 $F$ 且倾斜角为 $\frac{\pi}{6}$ 的直线 $l$ 与抛物线 $C$ 交于第一象
```

Here we know that `l` 's declaration covers span `直线 $l$` , so we starts from there. We just ignore the modifiers before `直线 $l$` , since they has nothing to do with the sentence.

Remember that the span for an assertion is the minimal part in natural language that can be translated to this assertion.

c. Queries

For queries, the span should include the natural language that represents the query term, along with the evidence that it is the query (e.g. 是什么, 为？).

Example:

双曲线C与椭圆$\frac{x^2}{36}+\frac{y^2}{16}=1$有相同的焦点,且$C$的渐近线为$x\pm\sqrt{3}y=0$,则该双曲线$C$的方程?

```
- queries:
Expression(C)
    [(109, 119)] ([双曲线$C$的方程?])
```

若双曲线的渐近线方程为$y=\pm3x$,它的一个焦点是$(\sqrt{10},0)$,则双曲线的标准方程是?

```
- queries:
Expression(C_1)
    [(41,55)] ([双曲线的标准方程是?])
```

In practice, we annotate the left and right index for each span. Annotators only need to select the spans in the questions and attach them to the corresponding assertions.

References

0 Default Individuals

0.1 Axis

Use `xAxis` to represent the X axis and `yAxis` to represent the Y axis.

No declarations!

Example:

点P在x轴上.

```
P: Point
PointOnCurve(P, xAxis) = True
```

0.2 Origin

Unfortunately, you need to declare a new variable in order to represent the origin point:

Example:

椭圆C的中心在原点上.

```
O: Origin
Center(C) = O
```

We need to take origin as a special concept. Sometimes the question won't mention what O is, but you are required to write the sentence `O: Origin`. Otherwise, the parser won't work.

Example:

已知P为椭圆$\frac{x^2}{8}+\frac{y^2}{2}=1$上的一个动点，$A(-2,1)$，$B(2,-1)$，设直线$AP$和$BP$分别与直线$x=4$交于$M$、$N$两点，若$\Delta ABP$与$\Delta MNP$的面积相等，则线段$OP$的长为?

```
O: Origin
...
```

Sometimes O may refer to other entities in the sentence. Then declare O as an instance of its true concept.

Example:

圆O与双曲线C相切于点$(5,0)$。

```
O: Circle
...
```

0.3 Constants

Feel free to use `pi` directly. Our system is quite familiar with this symbol.

0.4 rad, deg

The question may describe angles with units. You may use `applyUnit` to represent this:

Example:

$\angle ABC = 60^\circ$

```
A, B, C: Point
AngleOf(A,B,C) = ApplyUnit(60, degree)
```

$\angle ABC = \pi$

```
A, B, C: Point
AngleOf(A,B,C) = pi
```

where `degree` is a pre-defined individual. You should not declare it again.

0.5 Infinity

Mostly it only appears in the answers. Use `oo` to represent ∞.

Example:

$[3, \infty)$

```
[3, oo)
```

1 Entities with Properties

1.1 Expression

Basic expressions. Declare the variable, and write assertion(s) about its expression. Usually `Ellipse`, `Hyperbola`, `Parabola`, `Circle`, `Curve`, `Line` might have an expression.

If there are parameters in the expression, declare parameters like `a`, `b`, but not `x`, `y`. We never declare `x`, `y` since we think they are keywords.

If not explicitly mentioned, parameters are declared as `Number`.

Remember to write assertions about the constraints (if it exists).

Example:

已知椭圆C_1：$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1(a>b>0)$.

```
C1: Ellipse
a, b: Number
a > b
b > 0
Expression(C1) = (x^2/a^2 + y^2/b^2 = 1)
```

Be careful with the brackets `()` when using `=` !

Example:

已知双曲线C:$\frac{x^2}{2m^2}-\frac{y^2}{n^2}=1$.

✗:

```
Expression(C) = x^2/2*m^2 - y^2/n^2 = 1
```

✓:

```
Expression(C) = (x^2/(2*m^2) - y^2/n^2 = 1)
```

If there are constraints on `x` or `y`, use `And` to connect the constraints:

Example:

已知椭圆C：$\frac{x^2}{4}+\frac{5y^2}{4}=1(y\neq 0)$.

```
C: Ellipse
Expression(C) = And((x^2/4 + 5*y^2/4 = 1), Negation(y=0))
```

You may also use the syntatic sugar `&`:

Example:

已知椭圆$C:\frac{x^2}{4}+\frac{5y^2}{4}=1(y\neq 0)$.

```
C: Ellipse
Expression(C) = ((x^2/4 + 5*y^2/4 = 1) & Negation(y=0))
```

Be careful with the precedence! `&` > `,` > `=`. We recommend you use as much parenthesis as you can.

Example:

已知抛物线C:$y^2=-4x(y\geq 0)$.

✔:

```
Expression(C) = And((y^2=-4*x), (y>=0))
```

✔:

```
Expression(C) = ((y^2=-4*x) & (y>=0))
```

✘:

```
Expression(C) = And(y^2=-4*x, (y>=0))
```

✘:

```
Expression(C) = (y^2=-4*x & (y>=0))
```

✘:

```
Expression(C) = ((y^2=-4*x) & y>=0)
```

1.2 Coordinate

Tell the coordinate of a point.

Like the previous, if there are parameters in the expression, declare parameters like `a`, `b`, but not `x`, `y`. We never declare `x`, `y` since we think they are keywords.

If not explicitly mentioned, parameters are declared as `Number`.

Example:

点P的坐标为$(4, 3m)$.

```
P: Point
m: Number
Coordinate(P) = (4, 3*m)
```

1.3 LineSegment, Line, Vector

It is often to see 线段AB, 直线OP in the question texts. We use a constructor operator to represent them:

Example:

线段AB...

```
A, B: Point
LineSegmentOf(A, B)...
```

Similarly, we have `LineOf`, `VectorOf`. Also `TriangleOf`, `AngleOf`.

1.4 Distance, Length, Abs

These are the explanations for the three property operators:

`Distance`：点到点、点到直线、直线到直线的距离
`Length`：XX的长度（题面中出现“长度”）
`Abs`：|...| 中间是向量或线段

We only represent the question texts, so write sentences as it is.

Example:

线段AB的中点到y轴距离是3

```
Distance(MidPoint(LineSegmentOf(A, B)), yAxis) = 3
```

线段PQ长度的最小值为5

```
Min(Length(LineSegmentOf(P, Q))) = 5
```

$|AB| = 4$

```
Abs(LineSegmentOf(A, B)) = 4
```

$|\overrightarrow{AB}| = 4$

```
Abs(VectorOf(A, B)) = 4
```

$AB = 4$

```
LineSegmentOf(A, B) = 4
```

1.5 Vectors

There are two special things for vectors:

1. Use `DotProduct` to represent dot products.

Example:

$\overrightarrow{OA} \cdot \overrightarrow{OB} = 0$

```
DotProduct(VectorOf(O, A), VectorOf(O, B)) = 0
```

2. Use `0` itself to represent the zero vector($\overrightarrow{0}$, $\mathbf{0}$).

1.6 Angle

Simply use `AngleOf` to represent angles.

Example:

$\angle ABC = 60^\circ$

```
AngleOf(A,B,C) = ApplyUnit(60, degree)
```

$\angle ABC = \pi$

```
AngleOf(A,B,C) = pi
```

$\angle ABC = \angle BCD$

```
AngleOf(A,B,C) = AngleOf(B,C,D)
```

$\tan \angle ABC = 3$

```
Tan(AngleOf(A,B,C)) = 3
```

2 Set Domain

2.1 Multi-output Operators

Write a set when and only when there're multiple outputs.

Example:

已知直线L与抛物线交于A, B两点,与椭圆交于点C.

```
Intersection(L, E1) = {A, B}
Intersection(L, E2) = C
```

抛物线与直线L的交点在x轴上.

```
PointOnCurve(Intersection(E, L), xAxis) = True
```

椭圆C的焦点在y轴上

```
PointOnCurve(Focus(C), yAxis) = True
```

但不要在标注时直接表达:

✗:

```
PointOnCurve({A, B}, C) = True
```

✓:

```
PointOnCurve(A, C) = True
PointOnCurve(B, C) = True
```

2.2 Interval

Mostly it only appears in the answers. The same representation as in math. Use `+` to represent union.

Example:

直线l斜率的取值范围为$(2,3)$.

```
Range(Slope(l))=(2,3)
```

Further more:

x的取值范围为$(-\infty,-1]\cup(0,\infty)$.

```
Range(x)=(-oo,-1]+(0,oo)
```

2.3 ±

We use an individual `pm` to represent symbol $\pm$.

Example:

双曲线C的渐近线方程为$y=\pm\sqrt{3}x$.

```
Expression(Asymptote(C)) = Eq(y, pm*sqrt(3)*x)
```

The annotation rule is modified in this version. It is different from the 1st version.

2.4 OneOf

We use an psudeo operator `OneOf` to represent this relationship.

Example:

已知双曲线$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1(a>0,b>0)$的**一条**渐近线与直线$x+2y-1=0$垂直.

```
IsPerpendicular(OneOf(Asymptote(C)), l) = True
```

Example:

圆C经过双曲线的**一个**顶点和**一个**焦点.

```
C: Circle
E: Hyperbola
PointOnCurve(OneOf(Vertex(E)), C) = True
PointOnCurve(OneOf(Focus(E)), C) = True
```

The annotation rule is modified in this version. It is different from the 1st version.

Operator `OneOf` will finally turned into several assertions with operator `In`. We will not leave `OneOf` to the reasoning stage since it will cause induction problems.

Problematic questions: 263, 137, 132, 43, 23, 223, 49, 293,

3 Relationships

3.1 Tangent

We mainly have two kinds of relationships about tangent: '在...点处的切线', '过...点的切线'. The former one indicates that the point is on the curve, while the later one does not.

'在...点处的切线' uses `TangentOnPoint`, while '过...点的切线' uses `TangentOfPoint`.

Others just follow the operator definitions.

Example:

圆E与x轴相切

```
IsTangent(E, xAxis) = True
```

圆E与x轴相切于椭圆的右焦点F

```
F: Point
RightFocus(C) = F
TangentPoint(E, xAxis) = F
```

过F作圆O的两条切线,记切点为A、B

```
l1, l2: Line
TangentOfPoint(F, O) = {l1, l2}
A, B: Point
TangentPoint(l1, O) = A
TangentPoint(l2, O) = B
```

抛物线$y = x^2$在点P处的切线平行于直线$y = 4x - 5$.

```
C: Parabola
Expression(C) = ( y = x^2 )
D: Line
Expression(D) = ( y = 4*x - 5 )
P: Point
IsParallel(TangentOnPoint(P, C), D) = True
```

The operator names are modified in this version. It is different from the 1st version.

3.2 Chord

Chord is a relationship describing a line segment with two end points on a curve.

There are currently two operators related to chord: `IsChordOf`, `InterceptChord`.

Example:

直线$y = x$被曲线$2x^2 + y^2 = 2$截得的弦长为?

```
l: Line
C: Curve
Length(InterceptChord(l, C)) = ?
```

已知AB是过抛物线$y^2 = 2x$焦点的弦

```
A, B: Point
C: Parabola
IsChordOf(LineSegmentOf(A, B), C) = True
PointOnCurve(Focus(C), LineSegmentOf(A, B)) = True
```

If the question mentions 弦AB, then you are required to represent this chord relationship.

Example:

过点$M(1, 1)$ 作一条直线与椭圆$x^2/9 + y^2/4 = 1$相交于A、B两点,若M点恰好为弦AB的中点,则AB所在直线的方程为?

```
M: Point
Coordinate(M) = (1, 1)
l: Line
PointOnCurve(M, l) = True
C: Parabola
Expression(C) = (x^2/9 + y^2/4 = 1)
A, B: Point
Intersection(l, C) = {A, B}
IsChordOf(LineSegmentOf(A, B), C) = True
MidPoint(LineSegmentOf(A, B)) = M
Expression(OverlappingLine(LineSegmentOf(A, B))) = ?
```

3.3 Intercept

Intercept is a relationship between axises and lines. Axises include xAxis and yAxis. When using this operator, specify which axis to be intercepted.

Example:
直线l在y轴上的截距b

```
l: Line
b: Number
Intercept(l, yAxis) = b
```

4 Special Notice

4.1 No Rephrasing

Do NOT rephrase the sentence. Stick to the orginal expression.

Example:
抛物线 $y = (x - 2)^2 + 3$ 的顶点在直线l上。
✗:

```
E: Ellipse
Expression(E) = (y = (x-2)^2 + 3)
P: Point
P = Vertex(E)
l: Line
PointOnCurve(P, l) = True
```

✓:

```
E: Ellipse
Expression(E) = (y = (x-2)^2 + 3)
l: Line
PointOnCurve(Vertex(E), l) = True
```

Example:
抛物线 $y=(x-2)^2+3$ 的顶点P在直线l上。

✗:

```
E: Ellipse
Expression(E) = (y = (x-2)^2 + 3)
l: Line
PointOnCurve(Vertex(E), l) = True
```

✓:

```
E: Ellipse
Expression(E) = (y = (x-2)^2 + 3)
P: Point
P = Vertex(E)
l: Line
PointOnCurve(P, l) = True
```

4.2 When...

For 当...时, in most cases we treat them as facts, but sometimes we have to use special pseudo operators.

Example:
当 ΔFAB 的周长为3时，ΔFAB 的面积是?

```
Perimeter(TriangleOf(F, A, B)) = 3
...
```

Example:
当 ΔFAB 的周长最大时，ΔFAB 的面积是?

```
WhenMax(Perimeter(TriangleOf(F, A, B))) = True
...
```

Example:
当 ΔFAB 的周长最大时，ΔFAB 的面积是? 当 ΔFAB 的周长最小时呢?

```
无法标注，原因：其他
```

The annotation rule is modified in this version. It is different from the 1st version.

4.3 Quantifiers

Sometimes we need quantifiers to represent the question. We treat them as 'cannot annotate' with reason '其他'.

Example:
对于抛物线$y^2 = 4x$上任意一点Q，点$P(a, 0)$都满足$|PQ| \geq |a|$，则a的取值范围是?

```
无法标注，原因：其他
```

4.4 Numbers

We declare numbers if and only if they exist in the text.

Example:
双曲线C的离心率为2.

```
C: Hyperbola
Eccentricity(C) = 2
```

双曲线C的离心率$e = 2$.

```
C: Hyperbola
e: Number
Eccentricity(C) = e
e = 2
```

Cannot Annotate

Instruction

What questions cannot get annotated?

0. Out of the question;
1. Questions that are lack of operators/concepts;
2. Questions that need rephrase to annotate;
3. Questions that you are not sure how to annotate (remember to write remarks);
4. ...

Categorization

We divide all questions that cannot be annotated into the following categories:

1. Lack of Concepts/Individuals/Opeartors;
2. Question type does not match (E.g. Multiple Choices, Picture involved, etc.);
3. Involving knowledge in other domains;
4. Question proposes new definitions;
5. Ambiguous question description;
6. Questions with facts omitted;
7. Problematic questions;
8. Others.

1. Lack of Concepts/Individuals/Opeartors （算子缺失)

There are some Concepts/Individuals/Opeartors that is not included in lookup table.

2. Question type does not match （题型不符)

Some questions may be Multiple Choices. Some questions may have to use pictures provide important information. Some questions may ask students to choose all the correct statements from a list. etc. We do not consider all these question types.

3. Involving knowledge in other domains （知识点不符)

We only deal with questions that focus on the conic section part. Those requires knowledge about functions/polar-coordinates etc. are beyond our consideration.

Example:

已知抛物线C:$\begin{matrix} x=2t^2 \\ y=2t \end{matrix}$设$O$为坐标原点,点$M(x_0,y_0)$在$C$上运动,点$P(x,y)$是线段$OM$的中点,则点$P$的轨迹普通方程为?

无法标注，原因：知识点不符

4. Question proposes new definitions （新定义问题）

Some questions define some new stuffs. Though assertional logic is able to represent these questions, in this version we do not annotate these questions.

Example:

在平面直角坐标系xOy中,对于任意两点$P_1(x_1,y_1)$与$P_2(x_2,y_2)$的"非常距离"给出如下定义:若$|x_1-x_2|\geq|y_1-y_2|$,则点P_1与点P_2的"非常距离"为$|x_1-x_2|$,若$|x_1-x_2|<|y_1-y_2|$,则点P_1与点P_2的"非常距离"为$|y_1-y_2|$.已知C是直线$y=\frac{3}{4}x+3$上的一个动点,点D的坐标是(0,1),则点C与点D的"非常距离"的最小值是?

无法标注，原因：新定义问题

5. Ambiguous question description （题目歧义）

If you are not sure what is the meaning of the question (it is ambiguous), don't annotate it.

Example:

已知抛物线和双曲线都经过点$M\ (1,2)$，它们在 x 轴上有共同焦点，抛物线的顶点为坐标原点，则双曲线的标准方程是

无法标注，原因：题目歧义

To correctly annotate the sentence 它们在x轴上有共同焦点, we need to understand that the parabola has only one focus while the hyperbola has two foci. Not understanding this fact would lead to ambiguous annotations.

6. Questions with facts omitted （省略条件）

Some questions may omit facts that are obvious to human. But such things are not trival to machines.

Example:

已知双曲线的焦点在 x 轴上，且 $a+c=9$，$b=3$，则它的标准方程是?

无法标注，原因：省略条件

Here a、b、c are properties of the hyperbola. But the problem omits this fact.

7. Problematic questions （题目错误）

Some questions themselves might be problematic.

Example:

已知双曲线 C 经过点 $C(1，1)$，它的一条渐近线方程为 $y=\sqrt{3}x$. 则双曲线 C 的标准方程是

无法标注，原因：题目错误

C is hyperbola and point at the same time. This is not allowed.

8. Others （其他）

If you find the question cannot get annotated but none of the above reasons apply, select 其他. Write down the reasons in the `Remark` textarea.