

問題構造アノテーション仕様書

2013 年 5 月 21 日 版

目次

問題構造アノテーション仕様書.....	1
1 概要	1
2 文書情報.....	2
3 タイトル部分.....	2
4 問題	2
5 データ	3
5.1 テキストデータ	3
5.2 画像データ（写真、地図、概念図、グラフ等）	4
5.3 表データ	5
5.4 複数のデータのグループ	7
6 解答欄.....	8
7 選択肢.....	9
8 数式	12
9 テキストの装飾	12
10 空欄.....	13
11 クロスリファレンス	14
12 文字化け・欠落テキストの修正.....	14
13 レイアウト情報	14

1 概要

本仕様書は、「ロボットは東大に入れるか」大学入試センター試験関連オンラインタスクにおいて配布されるセンター試験問題データに対して付与されているアノテーションの仕様書です。

本仕様書で説明される主な内容は以下の通りです。

- 問題の表層的かつ直感的な構造に対するアノテーションの概要
- 参照関係等を記述するアノテーションの概要

この文書に出てくるアノテーション例はすべて、平成 22 年度大学入試センター試験の問題を利用しています。また、問題分類の詳細に関しては、「問題分類アノテーション仕様書」をご参照ください。

2 文書情報

試験問題の基本情報は、ルート要素の<exam>の以下の各属性によって記述しています。

<exam>

@source	問題の出題機関（大学入試センター）
@year	年度
@subject	科目名
@srcTxtURL	問題の URL（任意）
@range_of_options	（選択問題がある場合）選択問題の範囲（選択問題の id）
@num_of_options	（選択問題がある場合）選択する問題の数

3 タイトル部分

タイトル部分は以下の要素によって記述しています。

<title>	タイトル
<info>	解答番号の範囲や配点などの情報、「全問必答」などの指示。 （タイトルの直後だけでなく、問題の内部にも出現する。）

4 問題

問題は多くの場合、大きな問題がより小さな問いを含む入れ子構造になっています。本仕様では大小問わずすべての問題に対して<question>要素を使ってタグ付けしています。なおかつ、それ以上小さい問題を含まない問題を明示的に示すために、**minimal** 属性を用意しています。

<question> 問題の範囲を表す

@id	識別番号（文書内で固有）
@minimal	yes/no を取る。それ以上小問を含まない問題の場合のみ yes。
@answer_style	回答方式による問題の分類。（「 <u>問題分類アノテーション仕様書</u> 」を参照）
@answer_type	答えのタイプによる分類。 minimal 属性が yes の場合のみ指定。（別紙「 <u>問題分類アノテーション仕様書</u> 」を参照）
@knowledge_type	回答の際に必要な知識による分類。 minimal 属性が yes の場合のみ指定。（別紙「 <u>問題分類アノテーション仕様書</u> 」を参照）
@anscol	問題に対応する解答欄の id。複数ある場合は、「A1 A2」のように、スペースを挟んで複数指定。（ minimal 属性が yes の場合のみ指定。）

5 データ

解答者が解答を出すにあたって参考にする情報には、文章や図、表、およびそれらに対するラベル、説明文などが含まれます。ここではそのような情報全般を、<data>要素を利用してタグ付けしています。

<data> 文章や図、表などを含む、解答者が解答を出すにあたって参考にするデータ全般。
 @id 識別番号（文書内で固有）
 @type text, image, table, complex 等の値をとる。

<data>に共通して含まれる要素には、以下のものがあります。

<label> データにつけられたラベル（「表 1」「図 1」「A」「ア」など）
<caption> （主に表や図などに対する）説明文

語彙問題などで、解答選択のキーとして選択肢の直前に現れる語句なども<data>要素でタグ付けしています。

(1) 例：選択肢の直前に現れる語句

```
<question id="q07" minimal="no">
<label>B</label>  <instruction>次の問い（問 1～3）において、与えられた語と第一アクセント（第一強勢）
の位置が同じ語を、それぞれ 1～4 のうちから一つずつ選べ。</instruction>
<question id="q8" minimal="yes">
<label>問 1</label>
<data id="d01" type="text">damage</data>
<answer_column id="a05">5</answer_column>
<choices anscol="a05">
<choice><cNum>1</cNum> convince</choice>
<choice><cNum>2</cNum> effort</choice>
<choice><cNum>3</cNum> prefer</choice>
<choice><cNum>4</cNum> throughout</choice>
</choices>
</question>
.....（中略）
</question>
```

5.1 テキストデータ

解答の参考になるデータが文章である場合、<data>要素は type 値として text を取ります。そのような<data>要素は以下の要素を含む場合があります。

<quote> 文章中に現れる引用。(任意)
<note> 注。(任意)
 @id 注の識別番号(文書内で固有)
<source> 文章の出典情報。(任意)

(2) テキストデータのアノテーション例

<data id="d01" type="text">

 <ref target="n01">注 1</ref>フロイトによれば、人間の自己愛は過去に三度ほど大きな痛手をこうむったことがあるという。...(中略)...まずは<ref target="n02">(注 2)</ref>「ヴェニスの商人」について語らなければならない。...(中略)...

 だが、<uText id="u01"><label>A</label>経済学という学問は、まさに、このヴェニスの商人を抹殺することから出発した</uText>。

<quote>年々の労働こそ、いずれの国においても、年々の生活のために消費されるあらゆる必需品と有用な物質を本源的に供給する基金であり、...(中略)...したものである。**</quote>**

 『国富論』の冒頭にあるこのアダム・スミスの言葉...(中略)...一国の富の増大のためには外国貿易からの利潤を貨幣のかたちで<uText id="u02"><label>(ア)</label>チク</uText>セキしなければならないとする、重商主義者に対する挑戦状にほかならない...(以下略)

<source>(岩井克人「資本主義と『人間』」による)**</source>**

 (注) **<note id="n01">**1 フロイト——オーストリアの精神医学者(1858～1939)。**</note>**

<note id="n02">2 「ヴェニスの商人」——シェークスピアの戯曲「ヴェニスの商人」を...(以下略) **</note>**
</data>

5.2 画像データ(写真、地図、概念図、グラフ等)

データの内容が写真、地図、概念図、グラフ等の画像である場合、<data>要素は type 値として image を取ります。そのような<data>要素には、以下の要素が含まれます。

 画像
 @src 画像のファイル名
 @comment コメント(任意)

(3) 画像データのアノテーション例。左は元の画像（地理 B、p.158 より）。

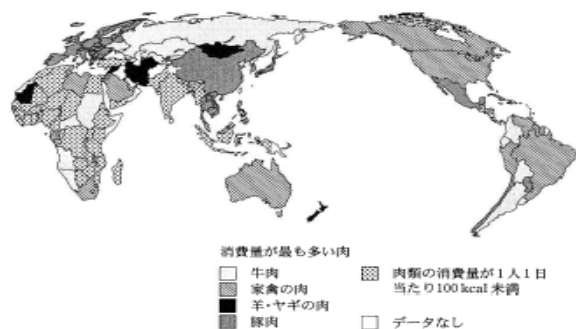


図 3

```
<data id="d01" type="image">

<caption>統計年次は2001～2003年の
いずれか。FAOの資料により作成。
</caption>
<label>図 3</label>
</data>
```

5.3 表データ

<data>要素のうち、表からなるものは type 値として table を取る。表の構造を記述する各要素は、ほぼ html における表の表現に準拠しています。（<tbl>は html の<table>、<row>は html の<tr>、<cell>は html の<td>に相当）

<tbl> 表。<row>、<cell>を含む。

<row> 表中の各々の行。<tbl>に含まれる。

<cell> 表中の各々のセル。<row>に含まれる。行内の<cell>の表示順は、左から右である。

（ただし、国語の場合は、右から左とする。）

@rowspan 複数のセルが横方向に連結されている場合。結合されているセルの数を指定

@colspan 複数のセルが縦方向に連結されている場合。結合されているセルの数を指定

(4) 表データのアノテーション例。上は元の表（地理 B, p.145 より）

表 1

	新庄市	金山町	舟形町	真室川町	最上町	大蔵村	鮭川村	戸沢村
15歳未満人口割合(%)	15.0	14.1	11.6	12.4	13.3	13.3	13.4	13.1
人口密度(人/km ²)	182.5	43.0	56.0	26.9	32.6	20.0	44.5	22.6
野菜の産出額(億円)	6.3	3.5	3.2	4.1	2.6	3.3	2.7	1.7

```

<data id="d01" type="table">
<label>表 1</label>
<tbl>
<row>
    <cell/>
    <cell >新庄市</cell>
    <cell >金山町</cell>
    <cell >舟形町</cell>
    <cell >真室川町</cell>
    <cell >最上町</cell>
    <cell >大蔵村</cell>
    <cell >鮭川村</cell>
    <cell >戸沢村</cell>
</row>
<row>
    <cell >15 歳未満人口割合（%）</cell>
    <cell >15.0</cell>
    <cell >14.1</cell>
    <cell >11.6</cell>
    <cell >12.4</cell>
    <cell >13.3</cell>
    <cell >13.3</cell>
    <cel >13.4</cell>
    <cell >13.1</cell>
</row>
.....
</tbl>
</data>

```

世界史に出てくる年表も、表データとしてタグ付けしています。

(5) 表データのアノテーション例。左下は元の年表（世界史 B, p.44 より）

1917 年	ロシア十月（十一月）革命
a	
1947 年	コミンフォルム結成
b	
1964 年	フルシチョフ解任
c	
1989 年	ベルリンの壁開放
d	

```

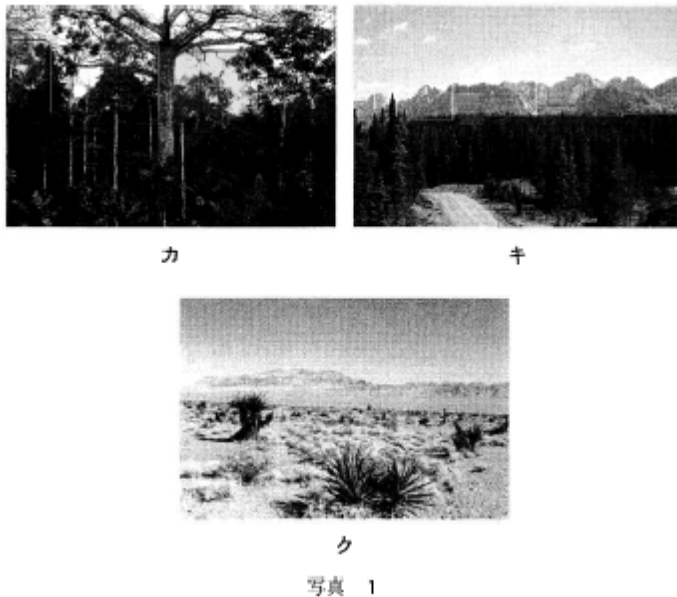
<data id="d07" type="table">
<tbl>
<row>
    <cell >1917 年</cell>
    <cell >ロシア十月（十一月）革命</cell>
</row>
<row>
    <cell rowspan="2">
        <blank id="b01"><label>a</label>
    </cell>
</row>
<row>
    <cell >1947 年</cell>
    <cell >コミンフォルム結成</cell>
</row>
<row>
    <cell rowspan="2">
        <blank id="b02"><label>b</label>
    </cell>
</row>
.....（以下略）
</tbl>
</data>

```

5.4 複数のデータのグループ

複数のデータ（それぞれラベルや説明文を持つ）のグループに、一つのラベルや説明文などが付いている場合は、<data>要素は type 値として complex を取ります。

(6) 複数の画像のグループに一つのラベルが付いている例。左は元の画像（地理 B、p.138 より）



```
<data id="d01" type="complex">
  <data id="d02" type="image">
    
    <label>カ</label></data>

  <data id="d03" type="image">
    
    <label>キ</label></data>

  <data id="d04" type="image">
    
    <label>ク</label></data>
  <label>写真 1</label>
</data>
```

6 解答欄

解答欄は、<ansColumn>でタグ付けしています。

<ansColumn>	解答欄
@id	識別番号（文書内で固有）

一つの問題に同じ番号を持った複数の解答欄が現れている場合は、いずれか一つが<ansColumn>によってタグ付けされています。

場合によっては、問題の範囲外に解答欄がある場合、明示的な解答欄がない場合、空欄の下に解答欄が埋め込まれる場合（例(7)）などもありますので、ご了承ください。

(7) 空欄の中に埋め込まれた解答欄

```
<question id="q05" minimal="no"><label>C</label><instruction>次の問い（問 1～3）において、それぞれ下の語句を並べ替えて空所を補い、文を完成させよ。解答は<ref target="b15">21</ref>～<ref target="b30">26</ref>に入れるものの番号のみを答えよ。ただし、文頭に置かれる語句も小文字で与えられている。</instruction>
```

```
<question id="q21" minimal="yes">
```

```
<label>問 1</label>
```

```
<data id="d07" type="text">Something you wrote in an email upset your friend. You could ask what had upset her by saying: Could you <blank id="b14"></blank><blank id="b15"><ansColumn id="a21"><label>21</label></ansColumn></blank><blank id="b16"></blank><blank id="b17"></blank><blank id="b18"><ansColumn id="a22"><label>22</label></ansColumn></blank><blank id="b19"></blank> with my email?
```

```
</data>
```

```
.....
```

```
</question>
```

```
.....
```

```
</question>
```

7 選択肢

選択肢のアノテーションに使う要素は以下の通りです。

<choices> 選択肢のまとまり。

@anscol 対応する解答欄の識別番号。複数ある場合は、「A1 A2」のように複数指定。

@comment コメント（任意）

<choice> 個々の選択肢。

@comment コメント（任意）

<cNum> 選択肢の番号（記号の場合も含む）。

選択肢がテキスト内に埋め込まれている場合、また表に埋め込まれている場合は、以下のようにタグ付けしています。

(8) テキストに埋め込まれた選択肢。上は元のテキスト（地理 B、p.146）

新田平岡集落の北を通る金山川に沿って①氾濫原^{はんらんげん}がみられる。共栄集落は②河岸段丘にあり、その集落の③約 600 m南には丘陵地もみられる。共栄集落と新田平岡集落との標高差は④約 30 mあるため、共栄集落は新田平岡集落に比べて水利条件は良くなかった。

```
<choices  anscol="031"> 新 田 平 岡 集 落 の 北 を 通 る 金 山 川 に 沿 っ て
<choice><cNum>1</cNum><uText id="u04">氾濫原</uText></choice>がみられる。共栄集落は
<choice><cNum>2</cNum><uText id="u05">河岸段丘</uText></choice>にあり、その集落の
<choice><cNum>3</cNum><uText id="u06">約 600m</uText></choice>南には丘陵地もみられ
る。共栄集落と新田平岡集落との標高差は<choice><cNum>4</cNum><uText id="u07">約
30m</uText></choice>あるため、共栄集落は新田平岡集落に比べて水利条件は良くなかった。
</choices>
```

(9) 表の中に埋め込まれた選択肢。左下は元の表。

	ア	イ
①	1.30	0.50
②	1.30	1.0
③	1.40	0.50
④	1.40	1.0
⑤	1.50	0.50
⑥	1.50	1.0

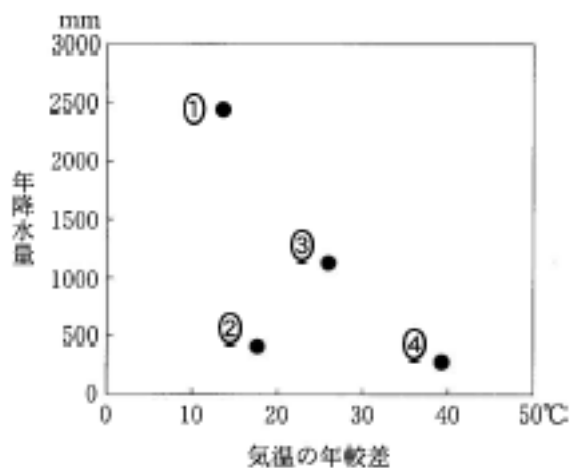
```

<choices anscol="a03">
<tbl>
<row>
    <cell/>
    <cell >
        <ref target="b05">ア</ref>
    </cell>
    <cell >
        <ref target="b07">イ</ref>
    </cell>
</row>
<row>
    <choice>
        <cell ><cNum>1</cNum></cell>
        <cell >1.30</cell>
        <cell >0.50</cell>
    </choice>
</row>
<row>
    <choice>
        <cell ><cNum>2</cNum></cell>
        <cell >1.30</cell>
        <cell >1.0</cell>
    </choice>
</row>
<choice>
    <cell ><cNum>3</cNum></cell>
    <cell >1.40</cell>
    <cell >0.50</cell>
</choice>
</row>
..... (以下略)
</tbl>
</choices>

```

選択肢が画像データに埋め込まれている場合は、要素を利用して、画像データの情報を<choices>要素の下に記述しています。画像の内部の個々の選択肢については現時点ではタグ付けをしていませんので、ご了承ください。

(10) 選択肢が画像データに埋め込まれている例



『理科年表 CD-ROM』により作成。

図 4

```
<choices anscol="a03">

</choices>
```

8 数式

現時点では、物理・化学の各試験問題ファイルにおいては、数式、座標、および変数の一部の情報が落ちており、復元はなされていません。本来数式、座標、変数が入る場所は、<formula>要素によって示しています。

<formula> 数式が欠落した部分
@comment コメント

数学においては、mathML を使用して数式を記述しています。

9 テキストの装飾

下線や傍線によって装飾されたテキストは<uText>、また下線や傍線はないが、ラベルの付いたテキストは<lText>でタグ付けしています。

<uText> 下線（傍線）付きテキスト

@id 識別番号（文書内で固有）

<lText> ラベル付きテキスト（下線（傍線）なし。「A」や「ア」などの記号が付いた短いテキスト）

@id 識別番号（文書内で固有）

<label> テキストにつけられたラベル

(11) 傍線付きテキスト、ラベル付きテキストへのアノテーション例 (国語、p.26)

殊方^あに なるまに おびたしう吹きまざりたる風の粉^{こな}れに、いと疾^{はや}う内裏^{うち}へ参^{まゐ}り給ひぬ。「今宵^{こんしやう}は中宮^{ちゆうきゆう}の御習直^{ごしうぢく}なりける
 が、下^{くだ}りさせ給ひけるまに、上^{うへ}は藤壺^{ふじやう}にわたらせ給ふ」と聞^{きこ}こゆれば、そなたさまへ参^{まゐ}り給ふに、立^たて蒔^まきなど、よろづの所^{ところ}あ
 らはに、例^{れい}ならず見^みわたされて、姫宮^{ひめみや}の御方^{ごかた}の御小壺^{ごこひ}の叢^{そう}に、童^{わう}べ下^{くだ}りて、虫屋^{むしや}ども手^てごと^{ごと}に持^もたり。御覧^{ごらん}するとて、二宮^{にみや}、
 御簾^{ごれん}を高^{たか}くもたげさせ給へるに、十一、二ばかりにやと見^みゆる御丈^{ごたけ}立ちにて、うつぶきて立^たち給へれば、前^{まへ}へ蒔^まき掛^かかれる御髮^{ごみづか}
 の削^けぎめふさやかに、絵^えに描^かきたらん心地^{こころ}して、まみ、額^{かぶ}、髪^{かみ}さし、か^(注五)の雪^{ゆき}の朝^{あさ}の御面影^{ごめんえき}なるものから、なほけしき異^いにて
 氣^き高^{たか}う、匂^{にお}ひも光^{ひかり}も頼^{たの}みなき御さまは、姫宮^{ひめみや}にこそはおはしますめれ。よろづのことに疑^{うたが}はず鎮^{しづ}まる御心^{ごこころ}も、たゞ今^{いま}はいかがはあ
 らん、深^{ふか}く心^{こころ}騒^{さわ}ぎして、おどろかれ給ふ。我が上の空^{そら}にもの憂^{うれ}く浮^うきたつ心^{こころ}は、この御さまなどを朝夕^{たふし}見奉^{みほう}らんには慰^{なぐさ}めなんか
 し、さりとして当時^(注四)、世^よの常^{じょう}に思^{おも}ひ寄^よるべき御年のほどならねど、ただまばり奉^{ほう}らまほしきに、「あはれ、雛屋^{ひいなや}に虫^{むし}のみよか
 し。一つにあらば、いかに嬌^{けう}しからん」とのたまへば、二宮^{にみや}、「あらわるや。昔^{むかし}や露^{つゆ}も入^いれさせ給はば、雛^{ひな}のため、いかにうつく
 しからん」と笑^{わら}ひ聞^{きこ}こえ給へば、げにと思^{おも}ひしるさまにて、まめだち給へる御まみのわたり、見る我^{われ}もうち笑^{わら}まれて、幾千代^{いくせんたい}
 まばるとも飽^あく世^よあるまじきに、おとなしき人^{ひと}参^{まゐ}りて引^ひき直^{ただ}しつれば、口惜^{くしやく}しうて歩^あみ過^すぎ給ふ。

A 宮城野^(注五)にまだうら若^{わか}き女郎花^{ぢやうらうが}移^{うつ}して見^みばやおのが垣根^{かきね}に

<data id="d11" type="text">

曉方に<uText id="u40"><label>a</label>なる</uText>まに、おびたしう吹きまさりたる風の紛れに、いと疾う内裏へ参り給ひぬ。「今宵は中宮の御宿直なりけるが、下りさせ給ひけるまに、上は藤壺にわたらせ給ふ」と聞こゆれば、....（中略）....かの雪の朝の御面影<uText id="u41"><label>b</label>なる</uText>ものから、なほけしき異にて気高う、...（中略）...世の常に思ひ寄るべき御年のほどならねど、<uText id="u42"><label>（ア）</label>ただまぼり奉らまほしきに</uText>、「あはれ、...（中略）...げにと思したるさまにて、<uText id="u43"><label>（イ）</label>まめだち給へる御まみのわたり</uText>、見る我もうち笑まれて、...（中略）...口惜しうて歩み過ぎ給ふ。

<|Text id="l01"><label>A</label> 宮城野にまだう
ら若き女郎花移して見ばやおのが垣根に</|Text>

.... (以下略)

</data>

10 空欄

空欄は<blank>要素でタグ付けしています。

<blank> 空欄

@id 識別番号（文書内で固有）

@digits 空欄に指定された桁数 (*理数系の問題で指定がある場合のみ)

<label> 空欄につけられたラベル

同じラベルを持つ空欄が複数現れており、同一のものを指示していると考えられる場合は、最初に出てきた空欄のみ<blank>でタグ付けし、残りの空欄は<ref>を付け、最初の空欄を参照するようにしています。

11 クロスリファレンス

クロスリファレンスには以下の要素を用いています。

<ref> 参照要素。

@target 参照先の要素の id の値をとる。複数指定の場合もある。

@comment コメント（任意）

<ref>は以下の要素を参照するテキストに付けられています。ただし、現時点ではあらゆる参照関係が記述されているとは限りませんので、ご了承願います。

◆ データ <data>

◆ 下線付きテキスト <uText>

◆ ラベル付きテキスト <lText>

◆ 注釈 <note>

◆ 空欄 <blank>

また、参照先の要素が画像内に埋め込まれている場合、現時点では target 属性は空にしており、comment 属性に「画像内」などのコメントを入れています。

12 文字化け・欠落テキストの修正

テキストファイル中の文字化けおよび欠落要素は、以下の要素を用いてタグ付けしています。

<garbled> 文字化け部分

@correction 正しい文字の記述

<missing> テキストが欠落した部分

@string 正しい文字の記述

13 レイアウト情報

 元テキストの改行位置を示す