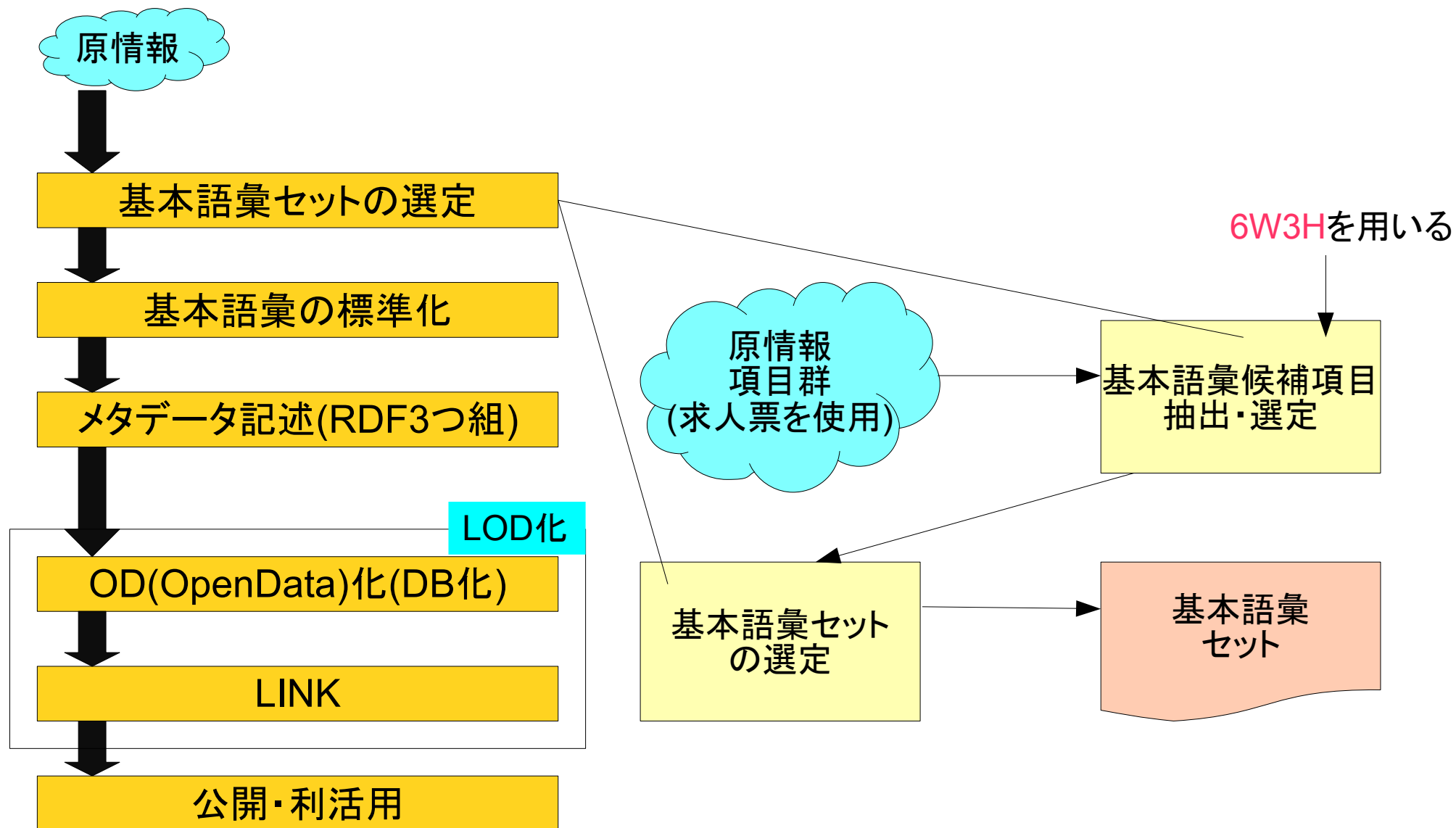


# 6W3H法による メタデータ化用基本語彙の選定方法

# 基本語彙セットの選定プロセス



例として求人情報を題材とした基本語彙を抽出

# 今回のチャレンジ内容

## ①基本語彙候補の抽出

- 5W1H(Who/When/Where/Why/What + How)  
+ 1W2H(Whom/How much/ How many)  
の分類軸を基本語彙候補の抽出に用いてみる

## ②基本語彙の選定

- ①の基本語彙候補から、樹形図を用いて基本語彙セットを選定してみる

# 求人情報のLOD化を題材に 基本語彙候補の抽出・選定を行ってみる

①原情報として大学に届いている求人票10社分をランダムに選択して用いる

- 企業毎にフォーマットは異なるが、記載項目は企業間での類似性は高い

②求人情報LOD化のメリット

- (1)求人情報をLOD化したシステムを利用することで、多くの企業が安価に求人情報を公開可能
- (2)他LODとのマッシュアップにより、多様な就職支援サイトの構築を容易にする

# 求人票記載項目の分類例(1/2)

6W3H要素	<p>           列挙した求人情報語彙            (基本語彙候補)         </p>
Who	<p>           求人先、本社所在地、<u>所在地</u>、<u>会社名</u>、<u>代表者</u>、<u>公開日</u>、<u>資本金</u>、<u>年商</u>、<u>売上高</u>、<u>従業員数</u>、<u>HP</u>、<u>URL</u>、<u>e-mail</u>、<u>TEL</u>、<u>株式</u>、<u>業種</u>、<u>事業内容</u> </p>
What	<p>           職務内容、<u>職種</u>、<u>職種名</u>、<u>役職</u>、<u>募集職種</u> </p>
When	<p>           フレックスタイム制、コアタイム、標準労働時間、就業時間、<u>勤務時間</u>、<u>休日</u>、<u>週休</u>、<u>休暇</u>、<u>休職</u>、<u>有給休暇</u>、<u>有給休暇数</u>、<u>夏季休暇</u>、<u>年末年始休暇</u>、<u>特別休暇</u>、<u>育児休職</u>/<u>育児休業</u>、<u>介護休職</u>/<u>介護休業</u>、<u>各種特別休暇</u>、<u>リフレッシュ休暇</u>、<u>永年勤続休暇</u> </p>
Where	<p>           支店等、支店、勤務先予定地、<u>勤務地</u>、<u>勤務地住所</u>、<u>事業所</u>、<u>営業所</u> </p>
Why	<p>           会社の特徴、<u>企業紹介</u>、特徴PR社風等、<u>募集理由</u> </p>

# 求人票記載項目の分類例(2/2)

6W3H要素	列挙した求人情報語彙 (語彙語彙候補)
Whom	<u>応募条件</u> 、 <u>応募要件</u> 、 <u>応募資格</u> 、 <u>短大・専門卒</u> 、 <u>大卒</u> 、 <u>募集対象</u> 、 <u>募集要項</u> 、 <u>資格等</u>
How much	<u>給与等</u> 、 <u>給与</u> 、 <u>基本給</u> 、 <u>初任給</u> 、 <u>賞与</u> 、 <u>昇給</u> 、 <u>給与改定</u> 、 <u>福利厚生</u> 、 <u>社会保険</u> 、 <u>加入保険等</u> 、 <u>労働組合</u> 、 <u>各種保険等</u> 、 <u>健康保険</u> 、 <u>厚生年金</u> 、 <u>雇用保険</u> 、 <u>労災保険</u> 、 <u>財形貯蓄</u> 、 <u>待遇</u> 、 <u>手当</u> 、 <u>諸手当</u> 、 <u>通勤手当</u> 、 <u>時間外手当</u> 、 <u>退職金</u> 、 <u>慶弔金</u> 、 <u>従業員持株制度</u> 、 <u>旅行など</u>
How many	<u>採用予定数</u> 、 <u>採用人数</u> 、 <u>募集人数</u> 、 <u>求人数</u>
How	<u>選考方法</u> 、 <u>選考フロー</u> 、 <u>試験内容</u> 、 <u>書類提出方法</u> 、 <u>提出書類</u> 、 <u>応募受付方法</u> 、 <u>応募方法</u> 、 <u>応募締切日</u> 、 <u>エントリー問い合わせ及び書類提出先</u>

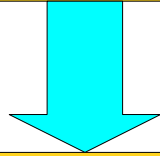
ページ4、5の下線付き用語を基本語彙セットとした

# 6W3Hの分類目安

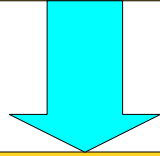
- Who
  - ・・・ 求人企業のプロフィール情報
- What
  - ・・・ 募集職種情報
- When
  - ・・・ 休暇、休日、労働時間情報
- Where
  - ・・・ 勤務地情報
- Why
  - ・・・ 企業のPR情報
- Whom
  - ・・・ 応募資格情報
- How much
  - ・・・ 賃金情報
- How many
  - ・・・ 採用数情報
- How
  - ・・・ 選考情報

# 求人情報メタデータ化用の語彙選定法

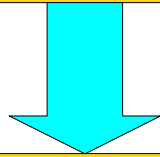
Step 1. G1～G9で列挙した求人票の基本語彙候補から  
同義語を代表用語に一本化



Step 2. 一本化された代表用語を樹形図化し、階層構造を可視化(図1)



Step 3. 用語の階層関係をチェックし、基本語彙を選定



Step 4. 6W3Hを用いて分類した全てのグループから選定した基本語彙を  
**基本語彙セット**として規定



# 例：樹形図化による基本語彙選定(図1)

①  
〈資本金〉 :capital

②  
〈休日〉 :holiday  
└─ 週休

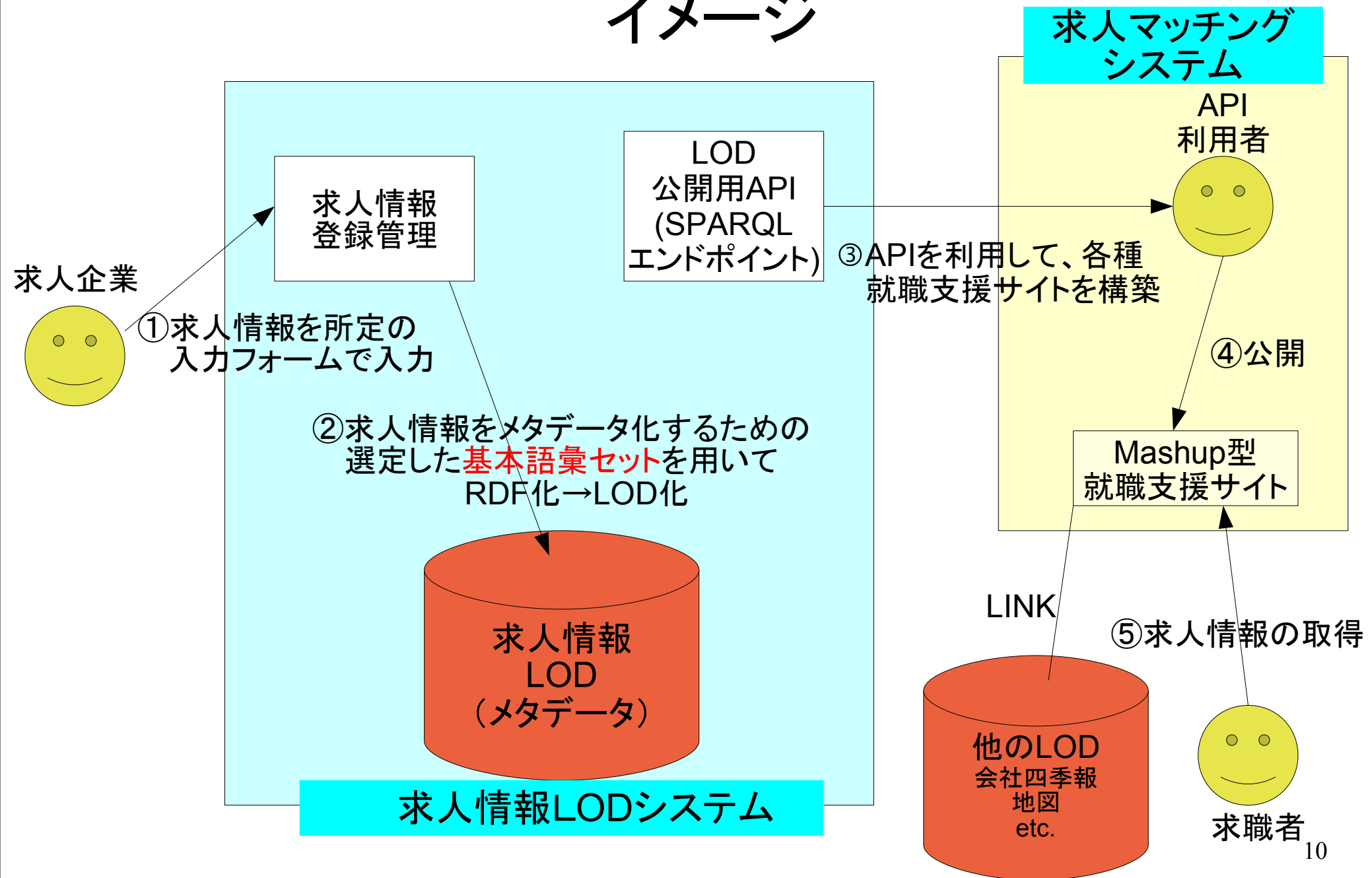
③  
〈休暇〉 :leave  
├─ 有給休暇  
├─ 夏季休暇  
├─ 年末年始休暇  
├─ 〈特別休暇〉 :special leave  
│ ├── 慶弔休暇  
│ ├── リフレッシュ休暇  
│ └─ 永年勤続休暇  
└─ 〈休職〉 :administrative leave  
 ├── 育児休職  
 └─ 介護休職

## 基本語彙の選定ルール

- ①子を持たない単一の用語
  - ②子を持つ親となる用語
  - ③最上位となる親用語、孫を持つ子用語
- 以上3つのパターンに適応可能な用語を基本語彙として選定

括弧<>で囲まれている用語を基本語彙として選定  
また、基本語彙名は英小文字アルファベット表現に変換して使用

# 求人情報のLOD化による活用 イメージ



# まとめ

- 求人情報のLOD化用基本語彙候補の抽出は6W3Hの分類軸を用いると可能で容易
- 基本語彙候補から基本語彙セットを選定する手段として樹形図は有力
- 今回の提案方法は、医療カルテ情報、不動産情報や、あるいは判例情報など他の分野のLOD化にも使えそうである