

Integrating and indexing variety of data using approximate matching and metric index

安達淳 高須淳宏 相澤彰子 深川大路 倉沢央

どんな研究？

重複情報を見つけ出すための要素技術として、
(1) 木構造データを対象とする近似マッチング

相澤 章

大規模データベース

ウェブ等の外部情報を利用したデータベースの再構築技術

The screenshot shows the iLinkage website interface. At the top, there's a navigation bar with the iLinkage logo and a 'About iLinkage' link. Below the navigation bar, there's a search bar and a list of search results. The first result is for 'Hsp90α', which is highlighted. The result details include the gene name, accession number, and a link to the full record. The interface is in Japanese.

- ✓ 数千万件規模のレコードを検索対象として、オンラインで同定結果を出力
- ✓ フォーマットを持たないテキストをデータベースのレコードに対応づける
- ✓ 入力テキスト全体ではなく、一部に照合するレコードが出力される
- ✓ 出力結果として得られるのは、類似レコードのランキングではなく、同一であると判定されたレコード

参照先

科研費報告書

論文書誌DB
CiNii
CrossRef

外部データベース

學術論文

学会・学術

研究者情

研究主屬性

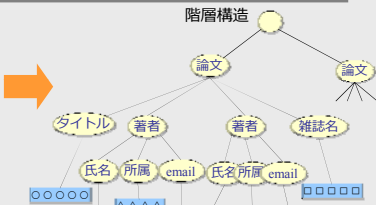
【作者】(日) 山本 浩一

7. 主題檢索

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

XML形式の文献データ

<論文>
 <タイトル>○○○○○○○○○<タイトル>
 <著者>
 <氏名>●●●●●<氏名>
 <所属>△△△△△<所属>
 <email>☆☆@☆☆☆<email>
 <著者>
 <著者>
 ...
 <著者>
 <雑誌名>□□□□<雑誌名>
 ...
 <論文>



距離空間を対象にした類似検索

```
graph TD
    Object[オブジェクト] --> Collection[集合データ]
    Object --> Image[画像]
    Collection --> List[List]
    Collection --> Set[Set]
    Collection --> Map[Map]
    List --> Vector[Vector]
    List --> ArrayList[ArrayList]
    ArrayList --> Tree[木構造データ]
    Tree --> A[A]
    Tree --> B[B]
    Tree --> C[C]
    Tree --> D[D]
    Tree --> E[E]
    Set --> HashTable[ハッシュテーブル]
    HashTable --> Array[Array]
    HashTable --> Matrix[Matrix]
```

© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd

Similar objects !

距離空間

あらかじめ指標とするオブジェクト (Pivot) からの距離で空間を分割し、距離の公理 (三角不等式など) を利用して検索処理の枝刈りする。

左の例では S_1 の部分空間を除外できる。

あらかじめ指標とするオブジェクト (Pivot) からの距離で空間を分割し、距離の公理 (三角不等式など) を利用して検索処理の枝刈りする。

左の例では S_1 の部分空間を除外できる。

分割マージンを最大化する
 ようなPivotを選択し、
 三角不等式による絞り込み範囲を向上させた。

検索時間の短縮（距離計算コストとディスクアクセスコストの削減）に役立つ。

分割マージンを最大化する
ようなPivotを選択
し、三角不等式による絞
り込み範囲を向上させ
た。

検索時間の短縮（距離計算コストとディスクアクセスコストの削減）に役に立つ。