

# iNLP Keywords

DEMO – July 14, 2010



Find the needle in the haystack



## An example

A researcher is interested in the heart rates of patients in medical reports. In particular, this researcher is interested in patients with heart rates of 100 or greater.

## Step 1

Find Heart Rates:

$\text{Exp} = \text{heart rate}$

Good, but what about pulse? It can be easily added:

$\text{Exp} = (\text{heart rate}|\text{pulse})$

**\* You can add HR, or other heart rate labels as needed. For this example, we'll use just heart rate and pulse.**

## Step 2

How do we capture the actual rate?

Exp = (heart rate|pulse)[^0-9,]\* ([0-9][0-9]\*)

The parenthesis represent “capture groups” and can capture specific sections of the regular expression.

## Step 3

Refine the expression so the heart rate number is within 20 characters of the words **heart rate** or **pulse**.

Exp: (heart rate|pulse)[^0-9,]{0,20} ([0-9][0-9]\*)

Now, limit to heart rates  $\geq 100$

Exp: (heart rate|pulse)[^0-9,]{0,20} ([1-9][0-9][0-9])